Intercreative Cinema:

Collaborative Expression with Digital Video

James Jung-Hoon Seo

S. B. Electrical Engineering and Computer Science
Massachusetts Institute of Technology
June 1994

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences
at the Massachusetts Institute of Technology

September 2001

Author
James Jung-Hoon Seo
Program in Media Arts and Sciences
August 24, 2001

Certified by
Glorianna Davenport
Principal Research Associate
Interactive Cinema, MIT Media Laboratory
Thesis Supervisor

Accepted by
Dr. Andrew B. Lippman
Chair, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Intercreative Cinema:

Collaborative Expression with Digital Video

James Jung-Hoon Seo

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on August 24, 2001
in partial fulfillment of the requirements for the degree of
Master of Science in Media Arts and Sciences

Abstract

Advances in technologies for digital video editing and streaming have lowered the barrier to entry for aspiring videomakers, and they provide an opportunity to expand the vocabulary for using and sharing video. Custom interfaces for editing and sharing video can suggest and support novel methods of collaborative production, cinematic narration, and casual dialogue with media. This thesis research presents Individeo, an online application for video browsing and editing, and explores how interface design can enable closer collaboration among online videographers. The thesis evaluates Individeo's custom interfaces through Honeymoon, an experimental collaborative video production, in which geographically separated videomakers attempt to build a cinematic narrative together through online collaboration.

Thesis Supervisor:
Glorianna Davenport
Title:
Principal Research Associate
Interactive Cinema, MIT Media Laboratory

Intercreative Cinema:

Collaborative Expression with Digital Video

James Jung-Hoon Seo

The following people served as readers for this thesis:

Reader
John Maeda
Sony Career Development Professor of Media Arts and Sciences
Associate Professor of Design and Computation
Program in Media Arts and Sciences

_____

Reader
Henry Jenkins
Director of Comparative Media Studies

_____

Credits

My parents and my sister, for your love, support and encouragement, all of which has helped me more than I can say in words. And for the Atari 800.

Glorianna Davenport, for giving me this opportunity, and for your vision, guidance and support during the last two years.

John Maeda, for giving me clarity and inspiration in times of confusion, and for pushing me to try harder.

Henry Jenkins, for all your insights and thoughtful comments, and for teaching me so much about film and media.

My colleagues at Interactive Cinema: Barbara Barry, Aisling Kelliher, Paul Nemirovsky, Pengkai Pan, Arnan (Roger) Sipitakiat, plus Yu Chen, Eric Gunther and Cathy Lin, and Ali Mazalek. Thanks for your support and insights during the past two years. Here's to the fun times in Dublin, Paris, and the DeCordova Museum. Cheers to the Shareable Media crew: Pengkai, Aisling, Cathy.

Jocelyn Lin, for her design and coding help. Thanks for being such a fun UROP to work with.

Phillip Tiongson, for your friendship, for shooting as the husband, and for your advice, enthusiasm and encouragement over the years. Our circle at IC is now complete.

Dong Shen and Tina Lee, for taking part in Honeymoon. I sincerely appreciate your time and your ideas.

David Cheng and Arun Tharuvai for helping to make Shareable Media a reality. David Crow, for working on Flights of Fantasy - thanks for your dedication.

David Kung, the original IC Super UROP, for all the great music and movies we shared, and for making possible the fantastic summer break in LA. (I can die now.)

Members of Interactive Cinema between 1991 and 1994, whom I had the pleasure of working with and learning from. Special thanks to Mark Halliday and Ryan Evans, who showed me the possibilities of multivariant movies. And to Amy Bruckman and Yun Kyung Choi, who first brought me to the Media Lab.

Tony Pulsone, for being a sturdy pillar of IC.

Linda Peterson, for your patience with my bouts of tardiness.

Eircom, for the sponsorship and the hospitality during my wonderful trip to Ireland.

Stan Huang and Steve Tom, who partly made it possible for me to come to the Media Lab. OMO and OJDK will always be close to my heart.

Unnam, for the cross-country road trip.
Alex, for the Hokkaido onsen (but where were the monkeys?) and the priceless VCDs.
Greg, for the IAP Hong Kong film fest.
Aimee, for all the volleyball that kept me sane and fit.

New Order and The Avalanches, for providing the soundtracks for my summer of 2001.

Contents

1.0 Introduction

1.1 Hypothesis

Making movies has never been easier: whether you are shooting a short film or recording a birthday party, the tools for shooting and editing movies in digital video have become increasingly accessible to both professionals and amateurs. Digital video cameras are growing steadily smaller and cheaper with each product cycle. The computer equipment needed for digitizing and editing video has declined in size and cost, and there are now video editing software targeted specifically at beginners. Companies tout home videomaking as a centerpiece of their marketing campaign for new computer products. Toy makers are building child-friendly video kits, encouraging kids to make masterpieces out of toy movie sets and characters. The media regularly hypes the ability of the amateur enthusiast to challenge Hollywood with homemade creations. Naturally, good ideas and hard work are still required to make movies worth watching, but the tools for making movies are within reach of a greater number of people than ever.

The World Wide Web offers an environment where the growing community of video users can share their creative output. A completed video can be placed on a networked server as a clip that can be downloaded or streamed by others on the net. The Web has become a content distribution medium with low cost and wide reach: although there is no guarantee that anyone will watch your birthday party video on the Web, anyone with Web access can at least get to see it. As the number of digital video users grows, a variety of video content has become available online. The spread of broadband networking enables a higher quality of video streaming, and improvements in streaming technologies make it possible to deliver video content to those without high-speed networking. The Web is also used for other purposes by videomakers: both professionals and amateurs use the discussion forums as a gathering place, coming together to discuss the latest tools and techniques.

There is an opportunity here to rethink and expand the way we make and share video, given the accessibility of tools and the connectivity of the Web. Traditionally, video editing is a solitary activity performed by an editor, who constructs the finished video out of the footage from the completed production. Today, "sharing" a video is limited to uploading the finished video clip and allowing others to watch it. In a networked video system, videographers and editors could come together during the editing process, rather than afterwards, and share their raw footage and ideas for how to edit it. Through a common storage database for shared source video, filmmakers could use each other's footage in addition to their own. They could collaborate with each other in near real-time to compose a video-based story together, comparing and discussing each other's footage and editing. Videomaking in a networked environment has the possibility of transforming editing into a less solitary and more communal endeavor. Accessible tools and widespread connectivity make this transformation a very real possibility.

The goal of this thesis is to propose a design for the software that can enable and support new forms of communal and collaborative video production. Existing software tools for video do not support collaboration among multiple filmmakers in an integrated fashion. As previously mentioned, video editing is considered a solitary activity: current video editing tools are designed for a single user who stores and edits all the footage locally. In a collaborative project using a networked video editing system, shared footage would be reused by multiple editors, who might choose to use the same video clip in different ways. For example, two editors might edit a scene differently with varying subsets of available footage; or one user might take clips from another user's home video, and put it in a different context for humorous effect. Traditional interfaces for video editing do not have a way of showing users exactly how multiple editors might be reusing a particular shared video footage. These interfaces reinforce the notion that video editing is an activity for a lone user.

It is proposed that the experience of collaborative video production can be enhanced by the use of custom visual interfaces. The interfaces would be tailored to the activity of collaborative editing with shared media. The user of the interfaces will be able to see clearly how shared video is being reused by different editors, and understand the different contexts into which shared media is placed. Each collaborator would be able to better comprehend how others are editing with the shared footage, and compare and discuss her work with collaborators. The hypothesis is that such interfaces can enhance collaborative authoring, by providing

information that is missing from traditional editing tools: contextual information about the shared media and the editing performed by each collaborator. The interfaces would also attempt to facilitate discussion and messaging among the online videographers.

1.2 Overview

This thesis explores the notion of intercreative cinema: filmmakers coming together on the Web, sharing their media, and collaborating with the aid of custom online tools. Tim Berners-Lee, the inventor of the World Wide Web, defines intercreativity as "the process of making things or solving problems together." He considers intercreativity a vital aspect of the future of the Web: "we ought to be able not only to find any kind of document on the Web, but also to create any kind of document, easily... We should be able not only to interact with other people, but to create with other people". [Berners-Lee 1999] To make this dream happen, Berners-Lee sees a need for online tools that can enable active co-construction, rather than just individual publishing or text communication. My research has focused on the design of such tools and their visual interfaces, within the domain of digital video production.

This thesis presents Individeo, a set of software interfaces for browsing and editing digital video. The Individeo browser provides a visualization of how multiple videomakers are reusing shared video in divergent ways. The browser uses a dynamic interface, based on generalized fisheye views, to scale and display a large number of video clips and sequences within a limited screen area. By showing the different ways a shared video clip has been edited within multiple sequences, the browser informs the user of the various contexts in which the shared video has been reused. The Individeo editor allows users to create video sequences composed of video clips and text inter-titles. Editing is performed through a storyboard interface, where sequences can be composed through the casual arrangement of video clips within an editing area.

This thesis includes an evaluation of Individeo through Honeymoon, an experimental video production involving the collaboration of two geographically separated videomakers. Starting with a shared outline for a narrative about a newlywed couple, two groups of directors and actors shot nearly simultaneously in Boston and New York, with a director-actor team in each city. The two teams used Individeo to share footage and edit sequences. The evaluation will consider how the Individeo interfaces can facilitate a novel collaborative production process. It should be noted that Honeymoon is a very particular form of collaboration, and there are other forms that Individeo might be well suited to. I decided to engage in the constrained production of Honeymoon in order to personally and closely evaluate the capabilities of Individeo; many of my observations from Honeymoon will be applicable to other types of video-based collaboration supported by Individeo.

1.3 Structure

Chapter 2.0, Extended examples, presents two detailed, specific scenarios of how Individeo might be used. The two scenarios are in the realms of home video and documentary filmmaking. Both examples involve collaboration in the form of sharing raw video and reediting shared footage.

Chapter 3.0, Theory and rationale, contains a review of relevant past research projects and collaborative productions. It describes films and online projects where collaborative construction is an important element in the creative process. An analysis of existing interfaces for video browsing and editing examines the current state of video-related software interfaces. Influential visual interfaces for displaying contextual information are presented, and relevant past projects from the Interactive Cinema group are highlighted.

Chapter 4.0, Design and implementation: Individeo, describes the centerpiece of this thesis, Individeo, an online application for video browsing and editing. The four design issues for Individeo's interfaces are outlined: shared contexts, scales of construction, casual editing, communication with media. Next, the features of Individeo browser and editor are described in detail. A brief description of the Shareable Media architecture is included.

Chapter 5.0, Evaluation: Honeymoon, discusses the experimental production project, which was used to evaluate the effectiveness of Individeo in an actual collaborative video production. While Individeo can be used within many different domains, I chose to evaluate it in the context of a constrained, personal production. The background for this decision is explained in the beginning of the chapter. The story outline and production framework for Honeymoon are presented, followed by specific examples of how Individeo was used during the production. The chapter ends with observations derived from the experience of the Honeymoon production.

Chapter 6.0, Conclusion, summarizes the two major components of this thesis: the development of Individeo and the production of Honeymoon. The thesis ends with final thoughts on networked video editing and sharing, and other applications for Individeo.

2.0 Extended examples

2.1 Home video

Bernard was as an exchange student from Manchester, England, who was spending a year in Boston at a local engineering university. An avid amateur video enthusiast, he had been shooting a video diary of his stay in America using his passport-sized digital video camera. Bernard was particularly excited about his first experience of the 4th of July celebrations. That night, he went to the Esplanade by the Charles River, and he was able to capture on tape the crowd of revellers, the symphony concert, and most importantly, the fantastic fireworks. The next day, Bernard decided he wanted to share his amazing footage with his friends back home. He digitized his footage, and made short video clips out of the best bits, including numerous shots of the blazing fireworks. He then logged onto the Shareable Media website. After uploading his clips to the online video database, he launched Individeo and retrieved his uploaded clips of the fireworks, which were now displayed in the Individeo browser. He picked out the ones he liked best and dragged the clips over to the Individeo editor. Within the editing area, he rearranged the clips to compose a video sequence. He positioned a clip of smaller fireworks in the top-left area of the editor to start the sequence, and ended with the clip of the biggest explosion in the bottom-right corner of the editor to finish the sequence with a bang. After adding an opening inter-title - "Summertime fireworks in Boston by Bernie" - to the beginning of the sequence, he uploaded the sequence to the Shareable Media website. He then emailed his friends about his latest cinematic masterpiece.

Peter in Manchester received Bernard's email, and visited the Shareable Media website. Using Individeo, Peter searched for clips of fireworks, and was able to see the clips that Bernard had uploaded. Using the browser, he selected and watched Bernard's sequence of fireworks clips. As he watched the succession of increasingly spectacular explosions, Peter thought of an idea for a humorous sequence that he could compose using Bernard's clips. First, he downloaded some clips for the film Gone With the Wind from a fan website. The clips were from the scene where Clark Gable and Vivien Leigh are about to embrace, and with each shot in the scene, the camera draws closer and closer to the faces of the stars, until the moment of the kiss, which is broken up by Leigh slapping Gable. Peter cut this clip into 3 separate shots, one for each level of close-up, and uploaded them to the Shareable Media website. In Individeo, he dragged these clips into the editor, along with several of Bernard's fireworks clips. By laying out the clips one after another in the editing area, Peter was able to quickly edit a sequence in which the fireworks clips are intercut with the romantic clips: the fireworks grow bigger and brighter as the two lovers come closer to each other, and the slap in the face is followed by the brightest explosion. Satisfied with his satirical sequence, Peter uploaded his creation. The Individeo browser was automatically updated, so that next to every instance of Bernard's fireworks clips that was used in Peter's sequence, the browser now displayed Peter's sequence automatically. Anyone watching Bernard's sequence would also see Peter's sequence at the same time, and would be able to instantly notice how Bernard's clips were intercut with other clips in Peter's sequence. The next time Bernard looked at his clips, he would have a pleasant surprise to watch and enjoy.

2.2 Documentary

Gillian was a filmmaker in New York, working on a documentary about the current state of electronic music. Over the past few months, she had compiled a large collection of video for her project: interviews with artists and fans, music videos, live concert footage. She had been using Individeo to organize and edit her footage on an ongoing basis. Each time she uploaded new clips to Shareable Media, she edited brand new sequences or modified existing ones; this allowed her to try out editing ideas during the production process, and store these ideas in the form of actual short sequences, rather than waiting until the end of production to begin editing from scratch. Yesterday, Gillian had shot an interview with a music critic who gave high marks to a band named Plaid. She digitized the interview footage into short soundbyte-sized clips and uploading them to Shareable Media. She then tried to edit a new sequence in Individeo with both old and new clips. Although she did not have any clips of Plaid themselves, she did find clips of other bands who were considered Plaid's musical peers. So she edited a sequence that intercut the latest interview clips with various music video clips. She laid out the clips in the Individeo editor: after a clip of the critic talking about the band's influence on the electronic music scene, she cut to a series of clips of music videos by various electronic bands. She ended the sequence with the critic musing fancifully about Plaid. She liked the sequence, but wished she could find actual video of the band in question. Before uploading the sequence, she added a text comment at the end of the sequence: "need footage of Plaid". This would serve as a reminder later when she saw the sequence again in the browser.

Stephen had just returned home from his favorite band Plaid's concert in San Francisco, excited about the video he was able to shoot on the sly from the upstairs balcony of the club. At the concert, his friend Rob had told him about an online documentary about electronic music. He logged onto Shareable Media, launched Individeo, and searched for clips associated with electronic music. He found a surprisingly large number of clips, and he noticed that many of the clips had been edited by Gillian. He watched several of Gillian's sequences, including the one about Plaid, which she had ended with a note about lacking actual footage of the band. Eager to share his new amateur video, Stephen digitized his concert footage and made small clips of what he considered as the show's highlights. He then edited a sequence, intercutting his concert clips with Gillian's interview clips. He used her sequence as the starting point, but created a new

sequence where he replaced the music video clips of other bands with his concert videos of Plaid. The music critic's words of praise in the last clip would be followed by a shot of Plaid and cheering fans. He also put together another sequence with just the concert footage, including some of his friends dancing. He emailed his friends to check out his sequences.

The next day, Gillian was elated to find that someone had uploaded clips of Plaid to Shareable Media. Watching the clips, she remembered another thread she was working on: the differences in the style of live concerts between rock and electronic music artists. She had collected some clips of live rock bands, thrashing their guitars and playing to the crowd. In contrast, the Plaid clips showed the duo calm and collected behind their laptops. She searched for and found the rock band clips using the Individeo browser, and she edited a sequence where she mixed clips of the two contrasting performance styles. She liked the jarring effect of combining the two sets of clips. Because the Individeo browser showed her all the sequences containing the Plaid concert clips, it was now easy for her to compare the many sequences created by herself and Stephen. She was able to see the different contexts that the clips were used in, which gave her an overview of the many directions her film could take.

3.0 Theory and rationale

This thesis is the result of my research in building tools for online collaborative expression with digital video. In order to set the theoretical background for my own ideas, this chapter presents a survey of previous works related to collaborative cinematic production, in the realms of traditional cinema, academic research and software industry. Because this thesis presents custom interfaces for video browsing and editing, an overview of current video software interfaces is included; because those custom interfaces stress the importance of providing users with contextual information about the presented data, this chapter presents visual interfaces for displaying contextual information which have influenced my own thinking. Relevant projects from the Interactive Cinema group at the MIT Media Laboratory are also discussed.

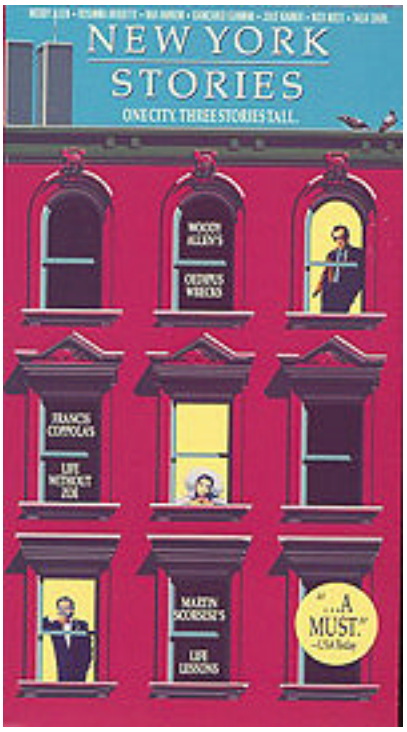
3.1 Collaborative production

Film production

The production of a feature film is a highly collaborative activity. Frequently, the achievements of a film are attributed to a single person, the director. Auteur theory in film studies attempts to analyze a set of films through the prism of the director's professional vision and personal life. Yet although certain directors may be considered singular visionaries, their films are shaped by the numerous hands involved with the production. Before shooting begins, the director might work with the screenwriter to flesh out the script, and with the art director to finalize the design of sets and costumes. During the shoot, the actors in front of the camera strive to give the desired performances, while the cinematographer behind the camera captures the images using the optimal lighting and camera setting for each given shot. During the post-production phase, after shooting is over, the final version of the film is constructed with the aid of the editor, the

soundtrack composer, and the special effects advisor. The vision for the film may originate from a lone filmmaker, but the implementation of that vision requires the coordinated efforts of numerous artists and craftspeople.

Some films use collaboration as a more explicit element of their formation, where the final result is composed of raw footage or edited segments that are contributed by multiple filmmakers. In omnibus films such as *New York Stories* [Allen, Coppola, Scorsese 1989] and *Four Rooms* [Anders, Rockwell, Rodriguez, Tarantino 1995], several directors agreed on an overall theme but produced their own discrete segments within the final film. The conceptual collaboration is defined up front, after which each filmmaker carries out independently her production process.

a. 

b. 

Figure 3-1. Omnibus films.
a. Poster for *New York Stories*: "One City. Three Stories Tall."
Directed by Woody Allen, Francis Coppola, Martin Scorsese.
b. The four directors of *Four Rooms*.
Left to right: Robert Rodriguez, Allison Anders, Quentin Tarantino, Alexandre Rockwell.

The Blair Witch Project [Myrick, Sanchez 1999] was almost entirely shot by the three main actors, sometimes using multiple cameras simultaneously; the directors limited their "directing" to leaving instructions for the actors / camerapeople to follow.

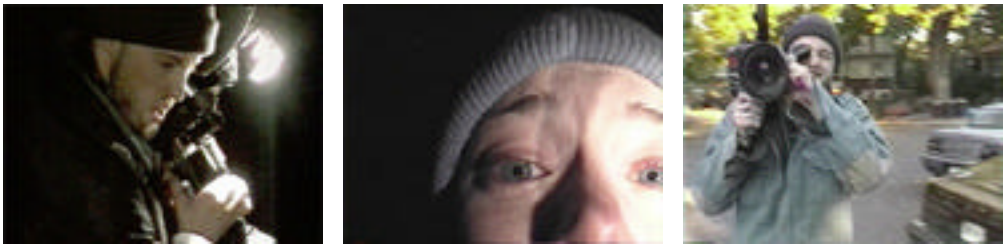


Figure 3-2. *The Blair Witch Project*.

This is What Democracy Looks Like [Freidberg, Rowley 2000] was shot by more than 100 cameras at different points in time during the WTO protests in Seattle, late 1999.

Figure 3-3. *This is What Democracy Looks Like*.

In the last two examples, there were only limited coordination and communication among the members of the creative team during the production. The respective directors waited until post-production to begin piecing together their films. *Timecode* [Figgis 2000] also featured multiple cameras shooting simultaneously, but the movement and filmed subject for each camera was carefully planned and coordinated in advance, using a shooting score borrowed from music composition.

a. 

b. 

c. 

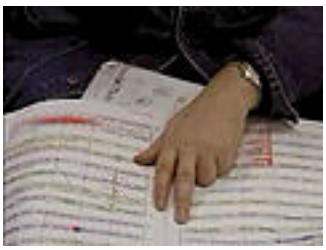
d. 

Figure 3-4. *Timecode*.
a. The film is composed of footage from four different cameras, which were shot simultaneously, and which are played back in four quadrants at the same time.
b. Director Mike Figgis, who was one of the four camerapeople.
c. d. The respective position, movement and captured subject of the four cameras (c) were carefully coordinated by using a shooting score (d), which borrows from music composition scores.

This thesis presents tools designed to support a particular type of cinematic collaboration, one that characterizes some of the above examples. Multiple videographers shoot and contribute their footage to a shared pool. All of the footage is made available to all of the contributors for viewing and editing, regardless of who produced which set of video. All the video content resides on a server, and any number of users can have constant access to all of the video clips.

Online collaboration

Artists in various fields are using the connectivity afforded by the Web to engage in online collaboration. The most basic form of such collaboration involves the exchange of raw media and unfinished pieces, which are transferred back and forth between the creative partners. In Coudal's Photoshop Tennis (http://www.coudal.com/tennis.html), an artist creates a digital image, which is sent to her collaborator; the second artist modifies the image using any software, and sends the transformed image back to the first

artist. The two artists continue the back-and-forth process as if in a tennis match or a conversation, adding additional layers or cutting out whole portions, reacting to each other's modification of the image.

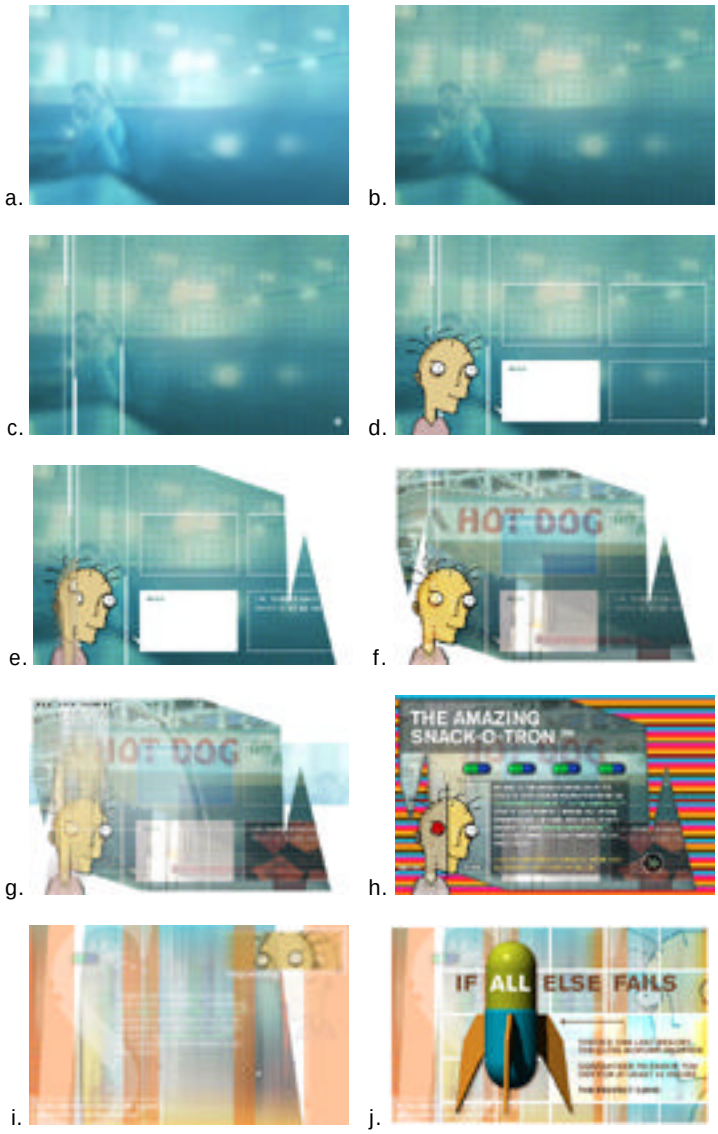


Figure 3-5. Photoshop Tennis, Coudal.
In the match shown here, Michelangelo Capraro and Michael Schmidt began with an original digital image (a).
The figures (a - j) show the progression of changes made by the two artists,
as they modified and sent the image back and forth.
The left column shows the work of Michelangelo Capraro; the right column, that of Michael Schmidt.

It should be noted that such a collaboration could have occurred offline, using regular mail to send storage media back and forth. However, because both the media and the tools for manipulating the media are digital, the Web becomes the natural transfer mechanism and exhibition space for this type of collaboration.

Communimage (http://www.communimage.ch/engl/) is an online collaborative project where the focus is on the collaborative evolution of a large body of media. The Communimage environment consists of a large 2D grid of digital images which collectively form a mosaic.

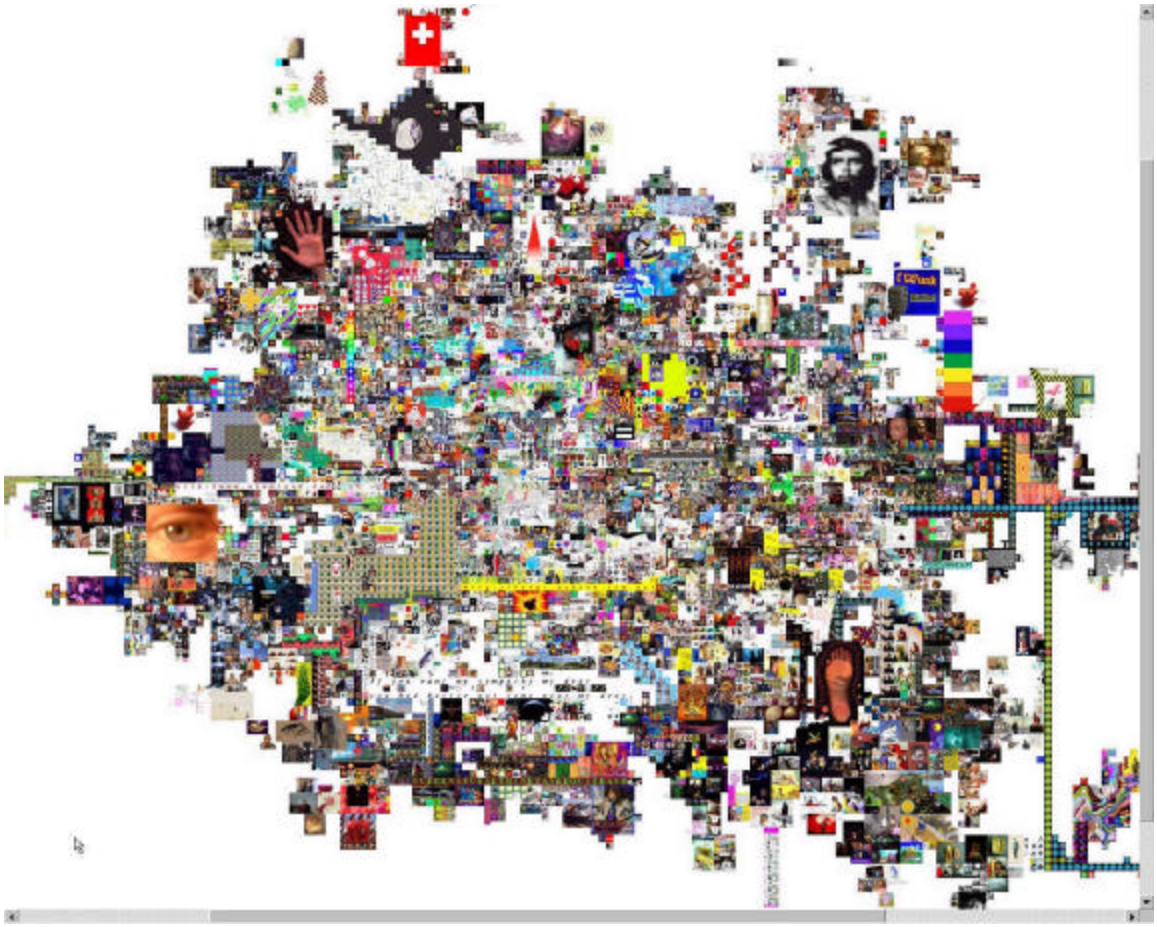
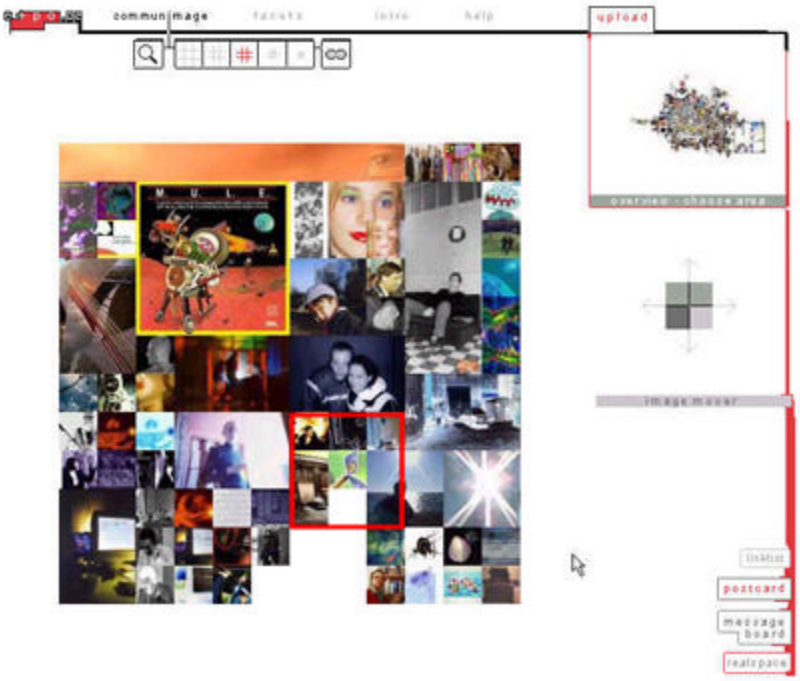
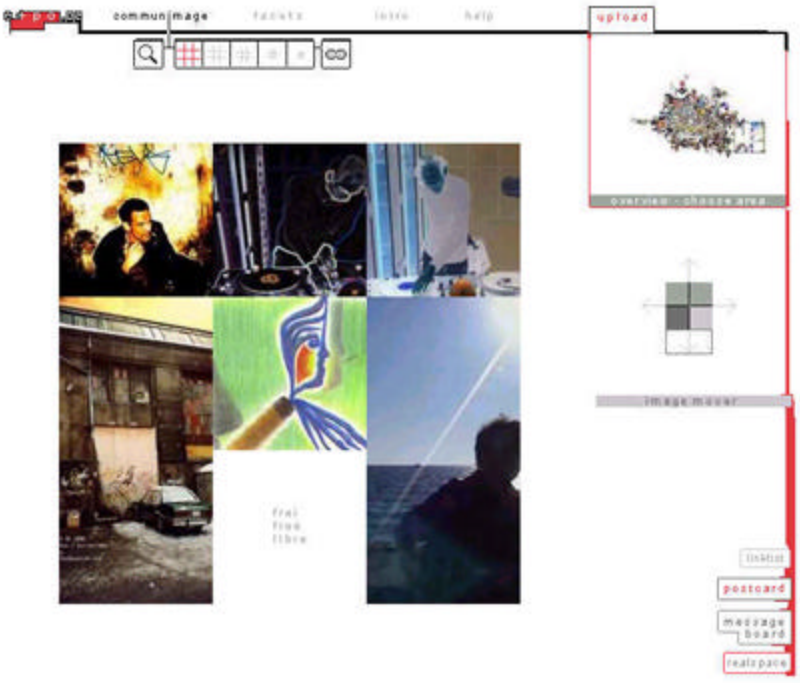


Figure 3-6. Portion of the Communimage mosaic of images.
Each square image in the grid has been uploaded by a user; the larger images
(such as the eye on the left part of the grid) are composed of many smaller images.

Users can inspect the grid to look at existing images, and upload a new square image to any open cell within the grid. The grid can be viewed at varying levels of zoom, so that the user can choose to see the overall collage of multiple images or study specific images on their own.



a.



b.

Figure 3-7. The Communimage interface for browsing the image grid.
The mosaic of images can be viewed at different zoom levels,
so that the user can either see the composite collage of multiple images (a)
or look at the individual images that make up the collage (b).

The simple interface of Communimage allows the users to juxtapose and play with images in different ways: they can build large-scale composite images using multiple grid cells; they can add parts of a bigger picture

but leave it incomplete, inviting others to fill in the blank spaces; they can upload a series of flipbook images as in a filmstrip; they can save an existing picture, modify it with image-editing software, and upload the changed picture next to the original image. In all cases, creating juxtapositions of images affects and modifies the perceived meanings of the images.

Figure 3-8. Four different sections of the Communimage image grid.

When I discovered Communimage, I was struck by its success in encouraging creative interaction and casual collaboration among a large number of users. I will return to Communimage in chapter 4 when I discuss the design and implementation of my own tool for collaborative construction.

3.2 Interfaces for video browsing and editing

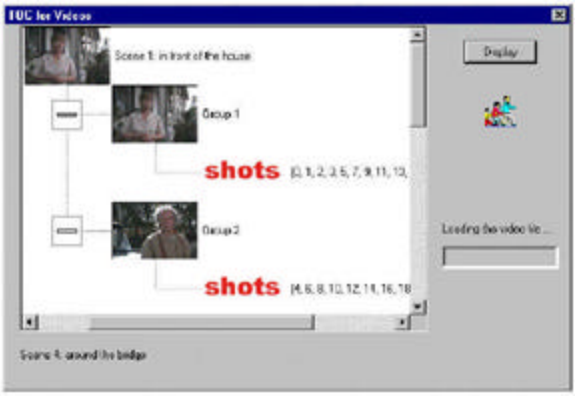
Video browsing interfaces

The widespread use of digital video has led to a growing body of research into the management of large video databases, one aspect of which is the development of tools for browsing video data. Although video browsing can be approached with general techniques for text or image retrieval, video data presents its own specific issues due to its temporal nature. The challenge lies in how to provide information about both the content of the video and the structure of how the video has been edited. Most video browsers represent the content of a video clip or stream using a key frame or a set of key frames selected from the video. The structure of an edited sequence of video clips is generally presented in one of 3 ways: timeline, hierarchy, and graph. [Aigrain 1996]

- The timeline is the simplest, where the key frames of each clip in a video sequence are aligned horizontally next to each other, in the same order as they appear in the temporal sequence. Almost all of today's video editing tools, described in detail later in this chapter, use this traditional representation.

Figure 3-9. Part of a timeline for video. From Apple iMovie.

- The hierarchical browsers provide several levels of information which the user can navigate: the top level may present key frames representing scenes, the next level may show sections of those scenes, and so on down to the level of shots. The multi-scale interface allows the user to analyze a video sequence and discover both the high-level structure and the shot-level details. [Yong 1999] [Chen 1998] [Zhong 1997]

a. 

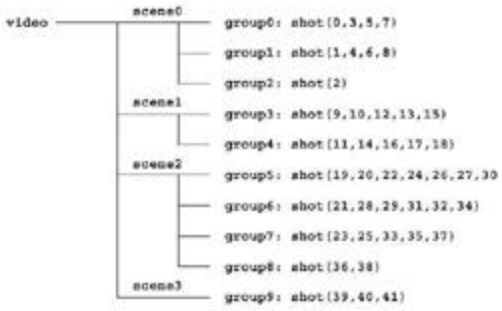
b. 

Figure 3-10. a. A hierarchical browser for video. [Yong 1999]
b. A sample hierarchy of a video segment, broken down into scenes and groups of shots. [Yong 1999]

- The graph is an alternative to the hierarchical model. The browser generates a directed graph for the given video sequence, in which each node is a cluster of visually similar shots. The browser represents each cluster node with a key frame from one of the cluster's shots. A link from one cluster to another cluster indicates that a shot from the first cluster is followed by a shot from the second cluster within the video sequence. A web of video clusters is formed by how the shots were edited in different sequences. [Yeo 1998] [Yeung 1995]

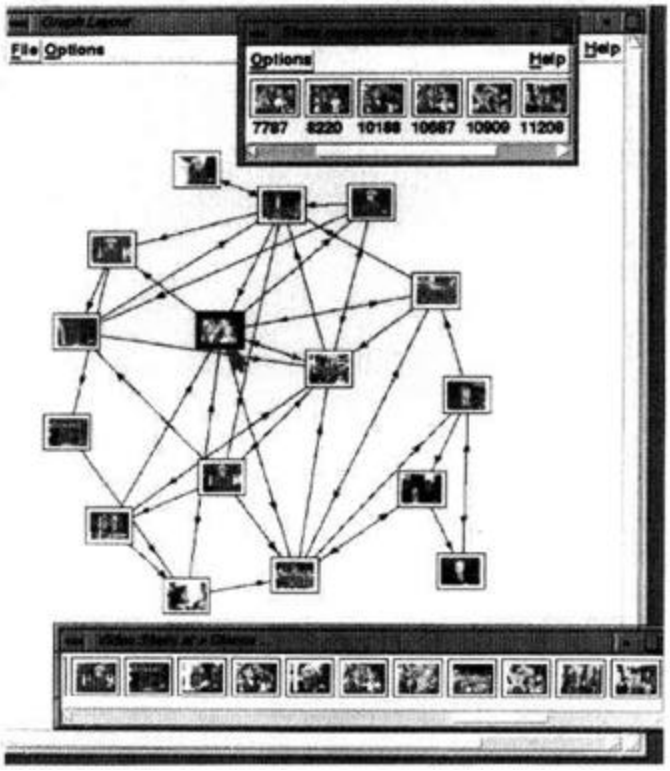


Figure 3-11. A graph representation of video. [Yeo 1998]

This thesis presents an alternative video browsing interface that uses a dynamic layout mechanism to present both the structure of a sequence and the links between sequences that contain the same clips. Later sections on the design of the browser will address its extensions beyond the three typical video browsing interfaces.

Video editing interfaces

Digital video editing systems have taken great strides since their first incarnations. The CMX-600, built by CBS and Memorex in 1971, is considered the first digital-disk based non-linear editing system. It was the first non-linear digital editing tool, because it freed video editors from having to lay down the video in the order they appeared in the final sequence. With a non-linear tool, the editors could start with any clip in a sequence, then add, modify or cut out clips at any time. The CMX-600 held 5 minutes of low-resolution black-and-white video on a disk drive the size of a refrigerator. Since the early 1990s, advances in computing hardware and video compression have led to the development of powerful editing systems that run on a desktop computer. The rising speed of processors and data transfer networks, coupled with the shrinking cost and size of storage devices, means that hours of high-quality digital video can now be stored and edited on a desktop computer. During the same period, Apple's QuickTime technology for manipulating time-based media such as video and audio led to the development of several non-linear video editing software utilizing QuickTime, such as Adobe Premiere and Avid Videoshop.

The interfaces for video editing have remained relatively consistent over the years. A recent and popular application for digital video editing is Apple Final Cut Pro.

Figure 3-12. Apple Final Cut Pro.

The 3 main components of the Final Cut Pro interface are the browser, the timeline, and the viewer.

- The browser displays video sequences and media files (video, audio, image) in the familiar tabular listing found in the Macintosh Finder or the Windows Explorer. Each clip is represented by an icon or key frame from the clip. The list of clips can be sorted using different fields such as name, duration, and frame size, and users can organize their media files by creating folders within the browser. The browser is often called the bin in other video editing tools, a carry-over term from the world of celluloid editing.

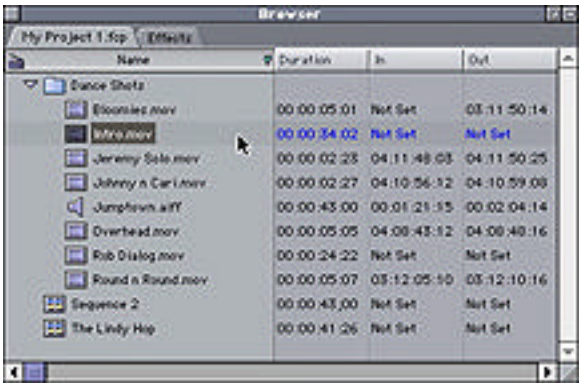


Figure 3-13. The browser in Apple Final Cut Pro.

- The timeline is where the user edits the sequence. It provides several tracks for video and audio. The user selects the desired clip from the browser, drags it into the timeline, and places the clip at the correct position within the sequence. Within the timeline, the user can reposition clips, trim the beginning and ending points of clips, and apply special effects to clips.

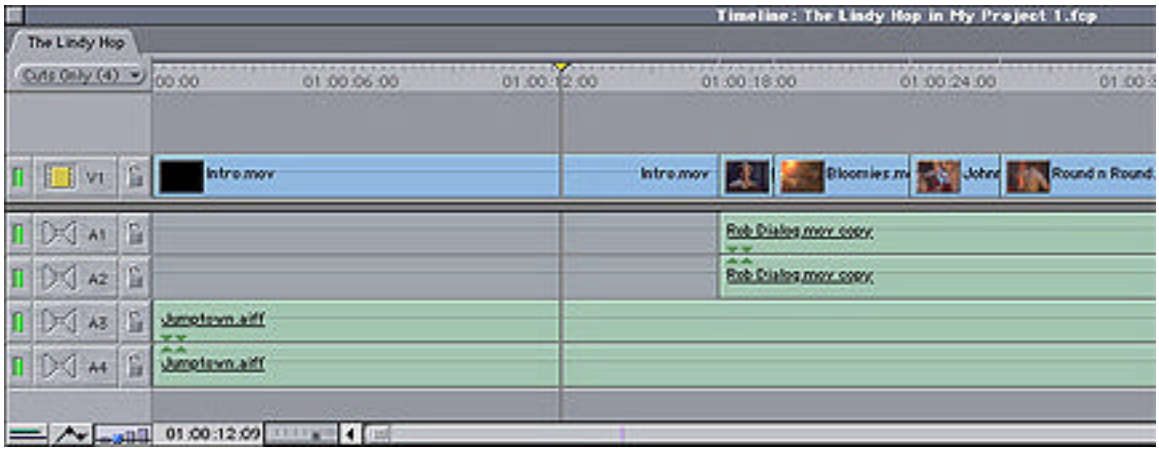


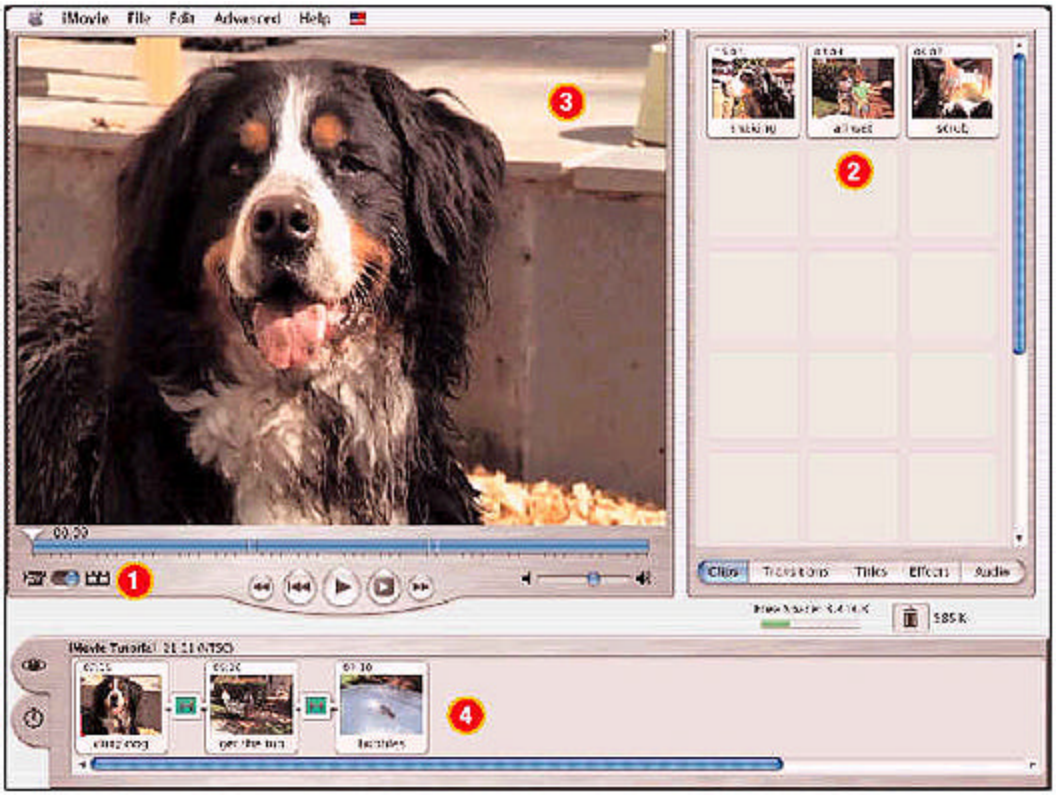
Figure 3-14. The timeline in Apple Final Cut Pro.

- The viewer plays back the video sequence or clip that is currently being edited.

Figure 3-15. The viewer in Apple Final Cut Pro.

These 3 components - the browser, the timeline, the viewer - could be found in video editing software in some form since the early days of desktop digital video editing. They can also be seen in most video editing tools available today, aimed at both beginners (Apple iMovie) and professionals (Media100 iFinish).

a.



b.

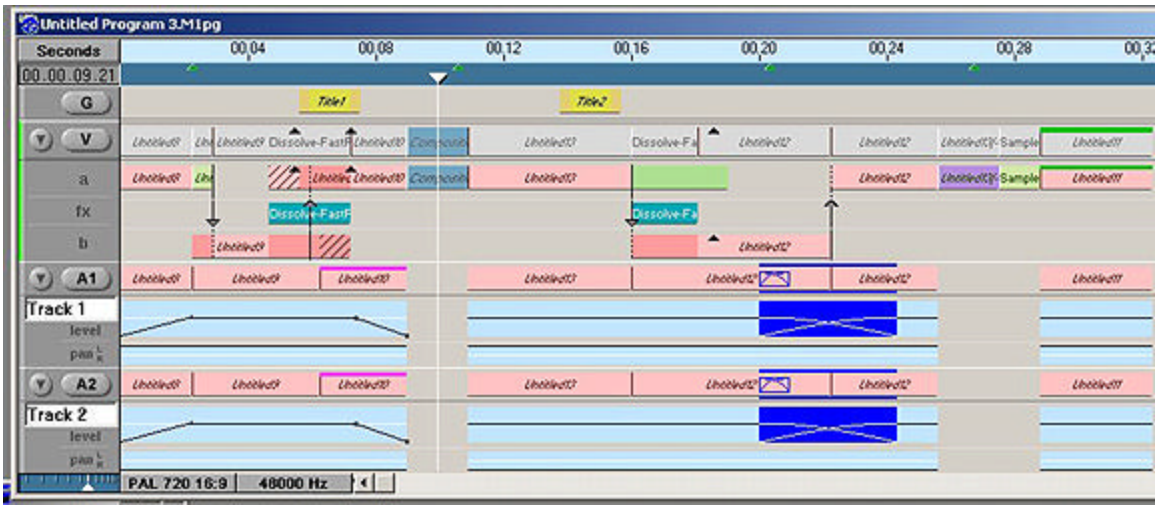


Figure 3-16. Common interface elements in current video editing tools.
a. Apple iMovie's browser (2), timeline (4) and viewer (1 and 3).
b. Sample timeline in Media100 iFinish.

There are a limited number of video editing tools built specifically for the Web. In most video-oriented websites, such as AtomFilms (http://www.atomfilms.com/) and Eveo (http://www.eveo.com/), users can submit a finished video clip for other viewers to watch online, but there is no support for editing online. D.FILM Moviemaker (http://w2.dfilm.com/index_moviemaker.html), Sony Screenblast (http://www.screenblast.com/) and Absolut Director (http://www.absolutdirector.com/absolut.director.asp) are three websites that allow users to edit together their own humorous animation, sci-fi video sequences and TV series trailers. The websites limit the visitors to a pre-defined set of clips to choose from; they cannot submit their own content for use with the website's editing tools. With GetMusic Videolab (http://videolab.getmusic.com/), users can upload their own media (video or still images) and edit sequences with various commercially available music tracks. All the online editing tools borrow the traditional interface elements for video editing (browser, timeline, player) with the exception of Videolab which uses a grid instead of a timeline.

a.

b.

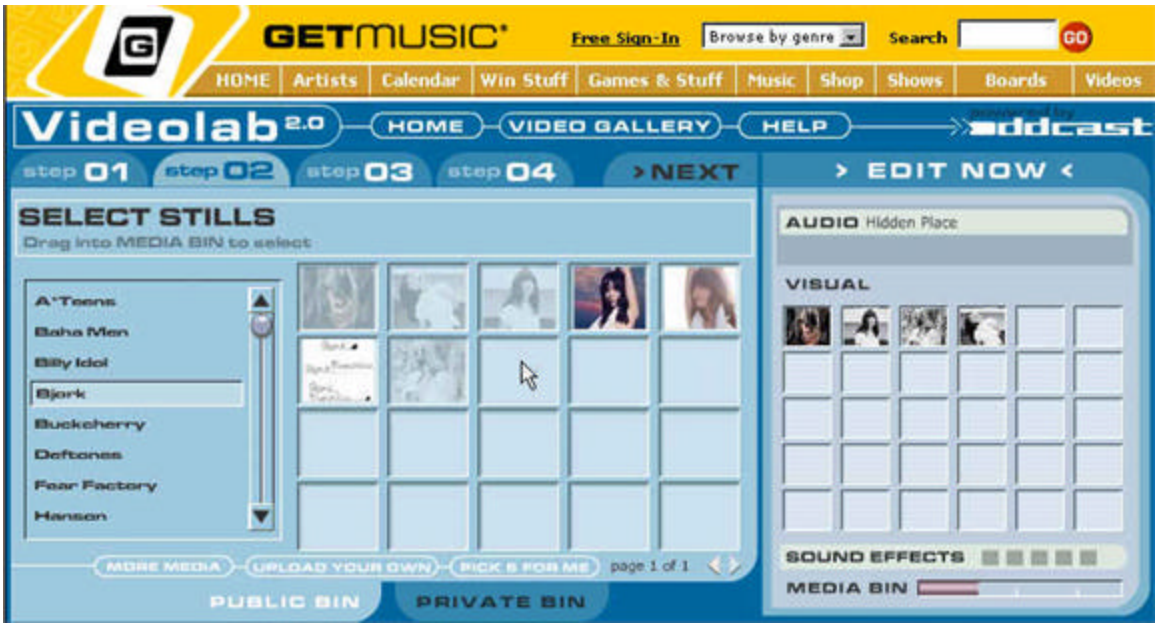


Figure 3-17. Online video editing interfaces. a. Absolut Director. b. GetMusic Videolab.

This thesis presents an alternative online video editing interface that moves away from both the timeline and the grid. A storyboard-like interface is proposed as a way of simplifying the editing experience.

3.3 Interfaces for visualizing contextual information

A central part of this thesis is the design of a visual interface that provides contextual information: the interface attempts to provide not only a set of information, but to visualize the relationship between a data point and the rest of the data set, so that the user is given a meaningful context for that particular data point. Here I describe some of the previous work in visual interfaces for displaying contextual information - projects that have influenced my own design ideas, presented in chapter 4.

Concordance software

Concordance is a tool used in literary and biblical analysis for analyzing the myriad contexts in which a particular word is used within a text. Concordance is defined as "an alphabetical index of the principal words in a book or the works of an author with their immediate contexts". For example, a concordance of a book with respect to the word "heart" will show all phrases containing the word "heart". The listing of the phrases will be organized to show the words that appear both before and after "heart". Looking at the list, the reader can obtain a quick overview of the different contexts in which the word "heart" has appeared in the book. The following two figures show two examples of concordance software for text.

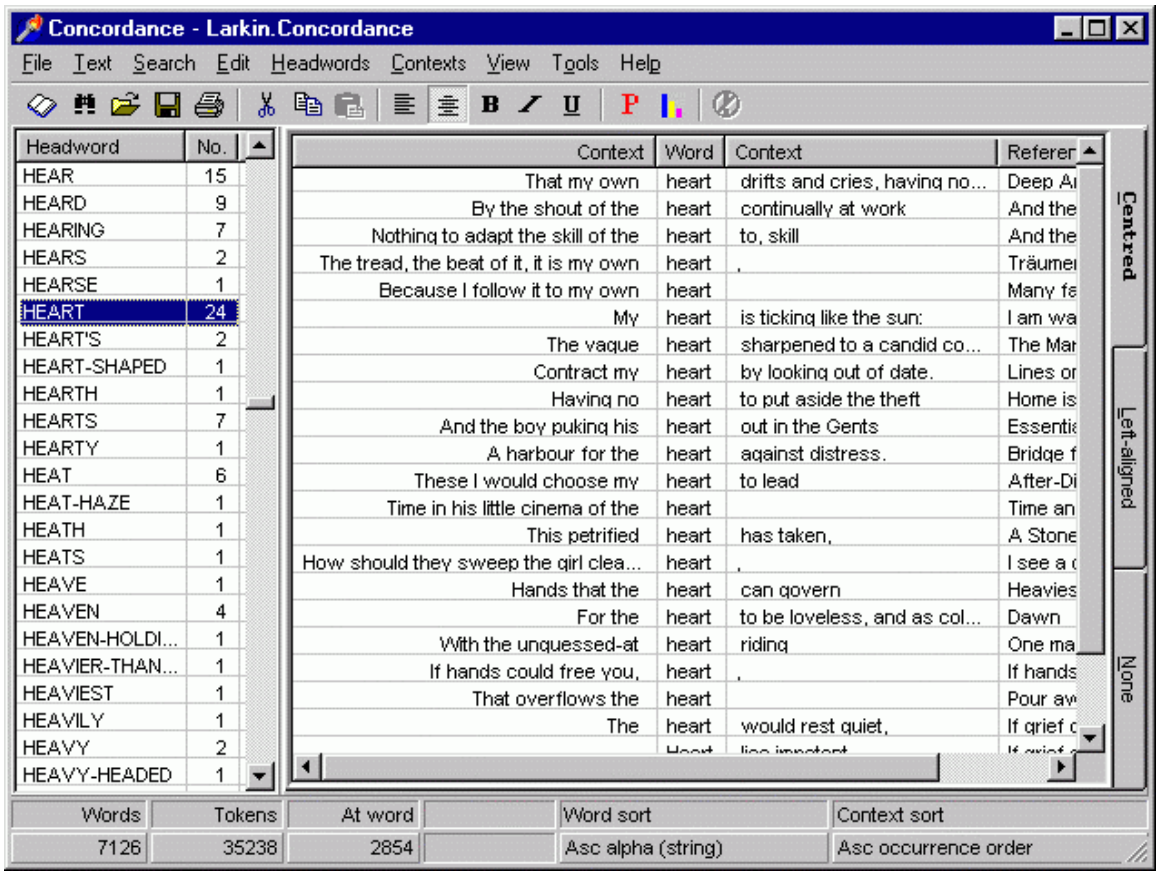


Figure 3-18. Concordance, a text analysis and concordancing software.
Analysis with respect to the word "heart" is shown above.
Developed by R. J. C. Watt.

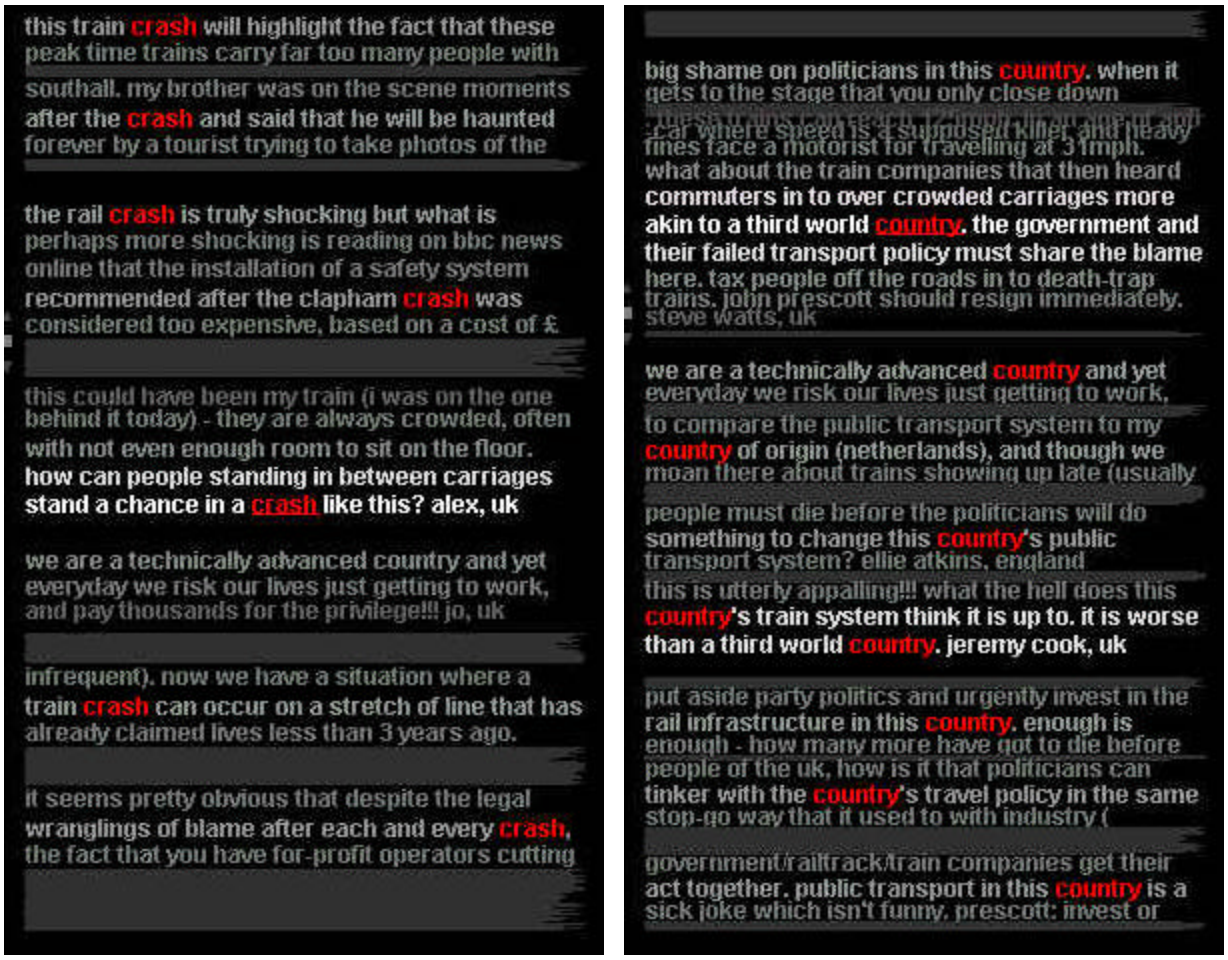


Figure 3-19. Jam! Comment Browser. (http://www.jamnewmedia.nl/commentbrowser/)
Analyses with respect to the words "crash" and "country" are shown above.
Developed by Michael Murtaugh.

Building a concordance of a text document is a useful method for analysis when one wishes to understand the various contexts in which a particular word has been used. Similarly, the Individeo browser, presented in chapter 4, creates a visual concordance by displaying all video sequences that contain a particular shared video clip. By visualizing the various contexts (sequences) a single clip has been used in, the browser builds a concordance of the video sequence database.

Generalized fisheye views

Fisheye views is a strategy for graphical display of information, formulated by George Furnas. It takes its name from the wide angle ("fisheye") lens that can show nearby objects in great detail while still showing all of the surroundings: it does so by presenting the other objects in successively less detail as they grow further from the current nearby point of interest. Saul Steinberg's famous illustration for New Yorker, *A View of the World from Ninth Avenue* (1976), is a famous illustration of the basic concept of fisheye views. In the painting, New York streets and buildings close by are depicted in great detail, while landmarks far away are shown in mere blobs and smudges.

Figure 3-20. *A View of the World from Ninth Avenue*, by Saul Steinberg.

The fisheye strategy attempts to view large information structures in a similar fashion, by providing a "balance of local detail and global context": "Local detail is needed for the local interactions with a structure... The global context is needed to tell the user what other parts of the structure exist and where they are." [Furnas 1986] Furnas defined a set of generalized concepts such as Degree of Interest (DOI) in order to implement the fisheye concept in software interfaces. Furnas also developed several sample applications to demonstrate the use of fisheye views, including the calendar shown below. The user of the

calendar can zoom into a particular day, and the rest of the month is scaled in size. The day of interest is shown in great detail, and the interface provides contextual information about the surrounding days.

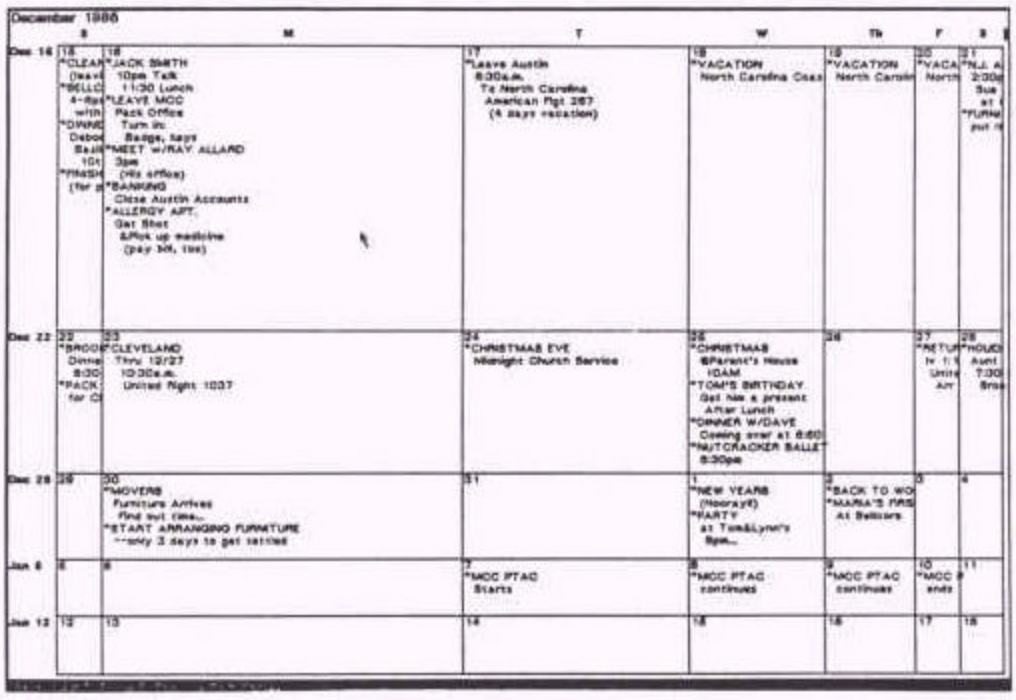


Figure 3-21. A fisheye views calendar. [Furnas 1986]

More recently, the Hyperbolic Browser developed at Xerox PARC extends the underlying theme of local detail and global context as outlined by Furnas. The Hyperbolic Browser uses a concentric display, in which a large hierarchy of information is laid out as nodes in a concentric graph. Concentric rings around the central root node indicate sub-layers of the hierarchy. The user can update the display in real-time by dragging any of the nodes, which updates the browser so that information about nearby levels is shown in detail, while other levels fade away or disappear.

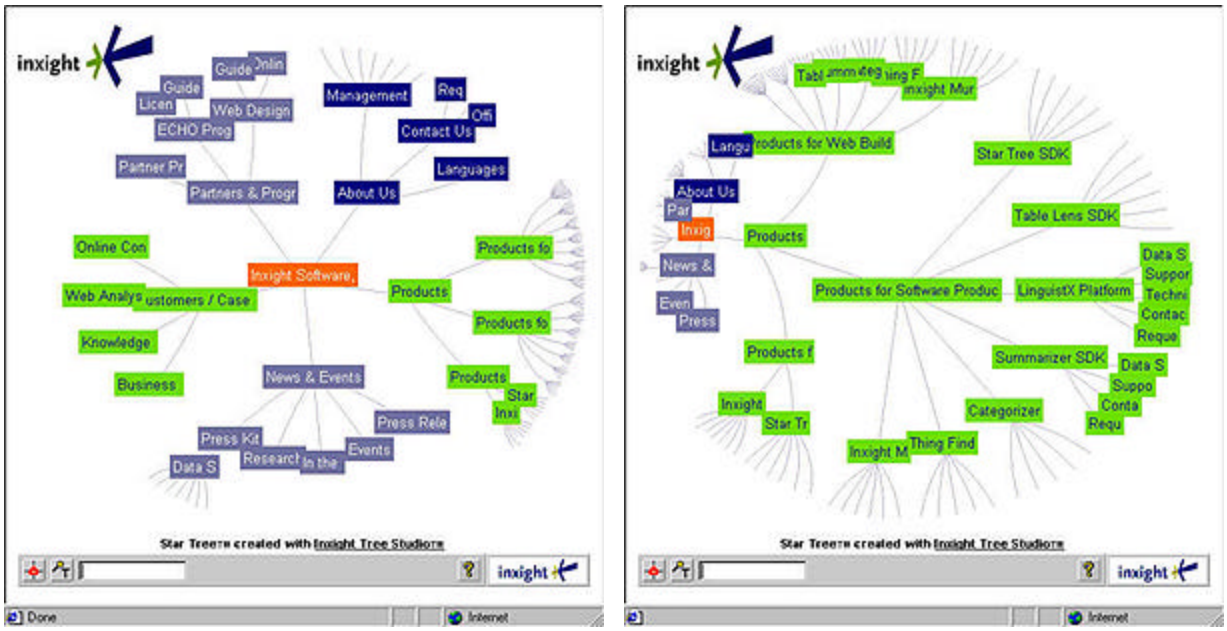


Figure 3-22. Inxight Star Tree, a descendant of Xerox PARC's Hyperbolic Browser. (http://www.inxight.com/)

Navigation of large databases through visual history

In his master's thesis, Axel Kilian developed various experiments that explored reactive space, user point of attention, and dynamic scaling of level of detail. Kilian's interfaces are beautiful examples of how the user's point of attention can be utilized within highly responsive visual interfaces. One of his experiments addresses specifically the navigation of large information structures.

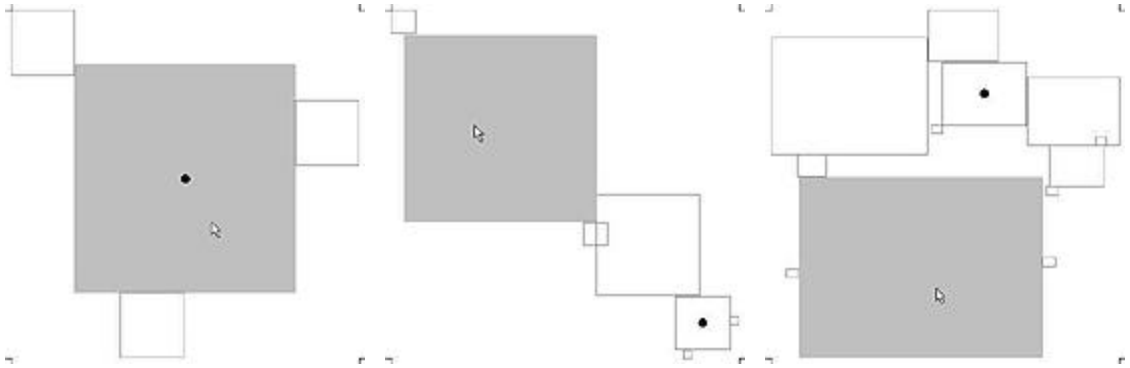


Figure 3-23. Experiment in navigation of large databases through visual history. [Kilian 2000] (http://destech.mit.edu/akilian/final/browse2/browse2.browse2.html)

In this experiment, a set of rectangular parent-child nodes are displayed in a structure where the child nodes are positioned and animated around the edges of the parent node. Each node can be expanded to reveal successive child nodes in the hierarchy, and all nodes remain onscreen once they have been expanded and made visible. The nodes maintain visual memory: the size of a node at any given time represents the amount of user attention on that node over time. This gives the user direct feedback about the navigation paths taken thus far, and the importance of each node as indicated by how much time was spent looking at that node. All child nodes are in constant animation around the edges of their respective parent nodes, rotating and adjusting its orientation relative to the center of the parent nodes.

I was inspired by the highly dynamic quality of Kilian's navigation interface. In my mind, he had taken a step beyond the fisheye-views interfaces of Furnas and others. In previous projects in fisheye views, the selection of a new focus and the subsequent transformation of the interface were carried out as discrete steps: user action would lead to a change in the display, which would become static again until further user action. In Kilian's experiment, the continual animation creates an interaction experience that is constantly revealing and engaging, and this is something I have attempted to capture with the Individeo browser. In the browser, moving the mouse continually will keep the overall visual form in constant animation, updating in real-time according to the mouse position. While the user can stop at any time, the highly dynamic interface aims to encourage the user to keep navigating along the structure of the visualized form, discovering new information. As chapter 4 will show, the Individeo browser is not as nearly volatile as Kilian's experiment. The Individeo browser becomes stable without constant and active user interaction, because the user needs to be able to stop the interface animation in order to study and select carefully from the displayed set of information.

3.4 Interactive Cinema

The Interactive Cinema (IC) group at the MIT Media Lab has been exploring the combination of cinematic narration and computation. Past research projects at IC have investigated new ways of construction, representation, annotation and interactive viewing of digital video. This thesis is an extension of one branch of research at IC: the development of computational tools for constructing and visualizing cinematic narratives. As such, this thesis draws from past IC experiments in the design of interfaces for video browsing and editing, and in the development of systems for supporting collaborative cinematic production.

Stratagraph by Thomas Aguierre Smith is a system that allows for the analysis of the significance of video content in terms of the different contexts of how it is used in editing. [Smith 1992] Instead of adding fixed annotations to a whole video clip, users can attach descriptions to any sequence of video frames within the clip. These descriptions have their own in and out points, and they can overlap to form layered streams of annotations for a given clip. When a sequence is edited with these clips, the title for the sequence becomes another type of annotation attached to the source clips. The aggregate set of temporal annotations (for a source clip and all the sequences that use the clip) can reveal the different contexts in which that video clip has been edited in: "the new context for an edited sequence effects the meaning of the frames that compose it. In this way, the chunks of video resource become embedded in a web of past uses." [Smith 1992]

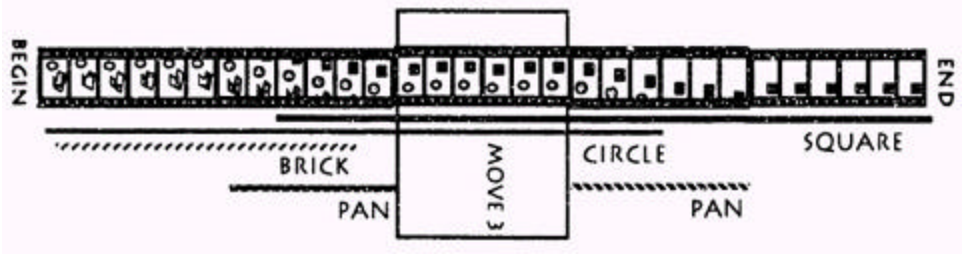


Figure 3-24. Describing video within Stratagraph. [Smith 1992]
The diagram illustrates how text annotations ("brick", "circle", "pan") can be specified for overlapping subsets of frames within the video segment.

Video Streamer and Collage by Edward Elliott introduce a dynamic 2-1/2-D stream representation of video and a collage-based editing space. [Elliott 1993] Most editing tools use a key frame or set of key frames to represent the video media. The Video Streamer displays video as a three-dimensional block: the video stream flows away in time and space, from the middle of the screen into the top-left corner. The successive rendering of the top and side of each frame in the video stream provides additional information about the temporal content, such as movement of camera or captured object. Users can make clips out of the Video Streamer by selecting portions of the stream. They can then drag the clips into the Collage space, where they can be laid out and played back in parallel, enabling "associative viewing" of multiple video streams.

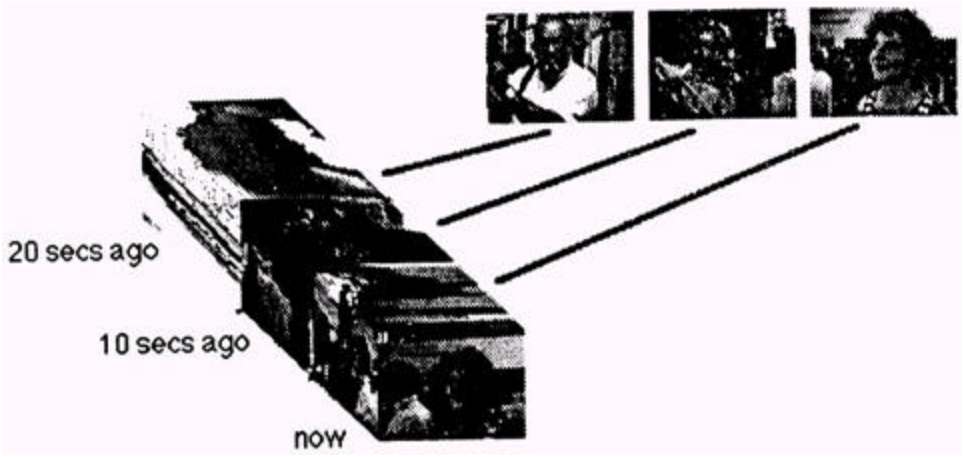


Figure 3-25. Video Streamer. [Elliott 1993]
A video stream flows away in 2-1/2-D space as it is played out.
Portions of the stream can be selected and dragged out as individual clips.
The figure shows three segments of the stream which correspond to three clips.

I-Views by Pengkai Pan is a system for authoring and publishing video-based narratives on the Web. [Pan 1999] The project fosters community building by allowing users to compare their narratives with each other, initiating dialogue based on their creative output. The system is composed of two types of tools: a sequencing tool that allowed the user to view, edit and publish video sequences, and a community-building tool where the user can compare different sequences and find other users with matching interests based on their stories. The goal of the project was to demonstrate a new story form named "Sharable Documentary" through the means of shared authorship, content and tools. I-Views is the predecessor to the Shareable Media system described in chapter 4.

Figure 3-26. The online sequencing tool in I-Views. [Pan 1999]
Different clips submitted by users can be arranged in order to create a sequence.

Flights of Fantasy, an art exhibition project led by IC, is an experiment in collaborative story construction. The project is an installation involving two rooms, where the visitors can engage in story making and viewing. In the story making room, the visitors can together construct stories in the form of video sequences using fragments of video. The collaborative interface for building sequences is specially designed for multiple users, who can choose the beginning and end points of the sequence within a grid of clips, and the desired types of video clips to add to a sequence. The physical interface is based on the children's pocket picture puzzle, with wooden markers and blocks; the visual interface (projected onto a wall) reflects the state of the physical interface, playing back the clips in the video sequences being constructed. Flights of Fantasy served as a localized experiment in collaborative cinematic production, where visitors are not passive viewers but are encouraged to both create their own stories as well as interact with other visitors.

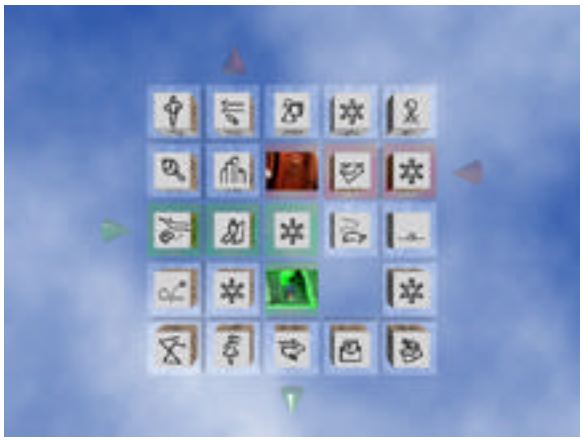
a.  b. 

Figure 3-27. Flights of Fantasy. The software interface for composing video sequences (a) can be controlled with the hardware interface by multiple users at the same time (b).

3.5 Summary

This chapter surveyed relevant past works in various domains: collaborative production in traditional film and content websites; software interfaces for video browsing and editing, and for displaying contextual information; and tools for video browsing and construction previously developed at Interactive Cinema. My research is an attempt to rethink and extend the interface design ideas embodied by these past works.

4. Design and implementation: Individeo

The power your (flattened) images have of being other than they are. The same image brought in by ten different routes will be a different image ten times. | *Robert Bresson, Notes on the Cinematographer*

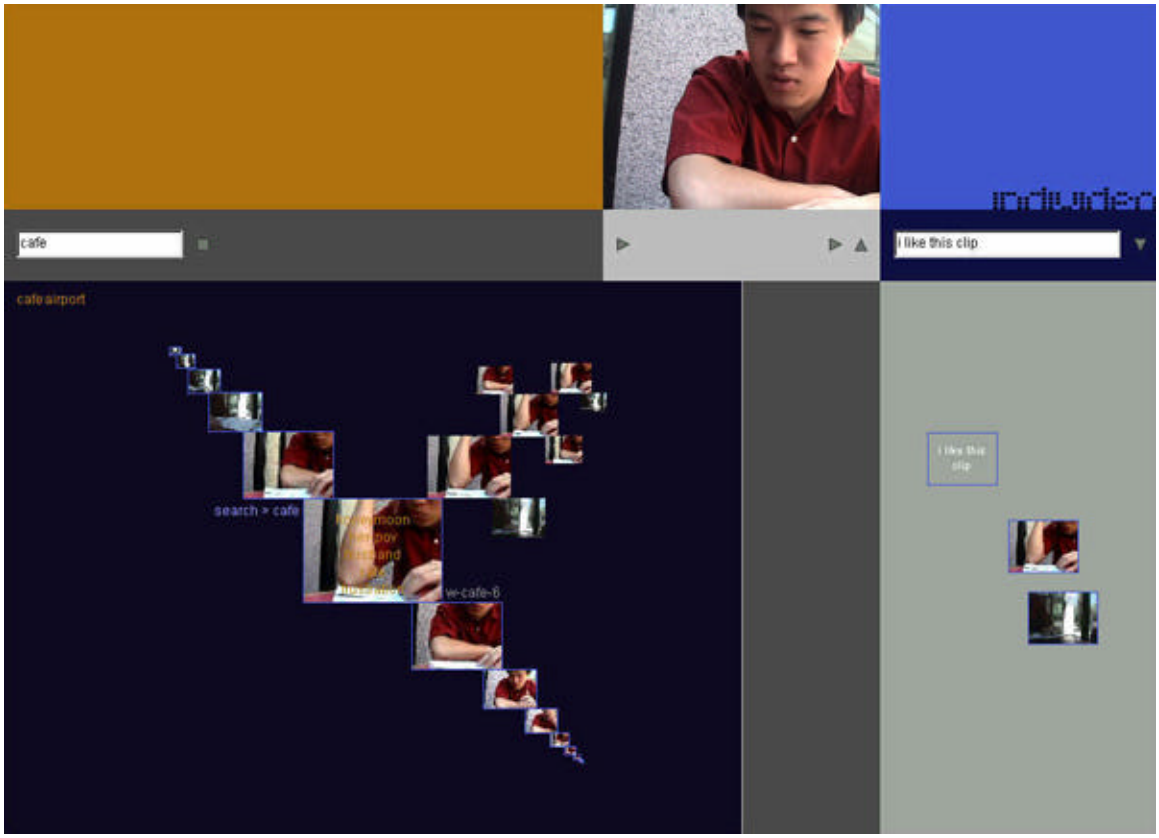


Figure 4-1. The Individeo interface.
All figures have been scaled down to save space.

Individeo is an application designed for online editing and sharing of digital video. Individeo users can browse and search a database of video clips, and edit and publish video sequences composed of those clips. The goal of this research is to ease and encourage collaborative construction with shared video, where multiple online users are editing a shared pool of raw video content. Individeo has two component interfaces, the browser and the editor. The Individeo browser supports the visual browsing and query of

shared video clips and edited sequences; the Individeo editor uses a storyboard interface for composing sequences with video and text clips. This chapter presents the design issues addressed by Individeo, and explains the functionality of the browsing and editing interfaces. It also describes the Shareable Media project which Individeo is a part of.

4.1 Design issues

While developing Individeo, I was looking for ways to engage and excite its users about sharing and reusing each other's content, and about communicating and collaborating with each other. I believed that developing the most sophisticated, feature-packed online video editor was not the answer. Instead, I focused on the question of how the software can make the individual user feel as if she were a part of the community built around media sharing and construction. Before discussing the features of Individeo, I present here the design issues considered during the development of Individeo. These issues also correspond to what I consider to be important aspects of collaborative media editing applications.

Shared contexts

Collaborative construction with shared media leads to the natural formation of shared context among the participants. If two editors are each editing a sequence with video clips, and they both use the same video clip in their sequences, the reuse of that clip forms the link between the two sequences, and provides a basis for comparison and discussion. Conversely, each new sequence that contains a particular clip results in an additional context associated with that clip. As the clip is juxtaposed among different clips in each new sequence, it acquires additional layers of meaning based on the context of the new sequence.

The interface for construction with shared media can visually provide contextual information, a valuable resource for collaborating users. If a user has contributed a video clip to the shared pool, she could later use the interface to see how other editors had used her clip. She may feel a closer connection to these

users, since she is able to understand how her contribution has affected their creative output. In an online setting, the user is also likely to log on regularly to find out how her clip has been reused since her last session. Similarly, if the same user selects some clips and builds a sequence, the tool could show her other sequences that make use of the same clips she had chosen. This would link her sequence with those of the other members of the creative community, and she could discover the variety and richness of the different ways in which her fellow editors have approached the same content. It has been noted that "the word context comes from the Latin cum (with) and texere (to weave) and etymologically suggests a process of weaving together." [Brown 2000] The Individeo browser attempts to make explicit this web of interconnections resulting from construction with shared media.

Scales of construction

When editing with shared media, each user can think about her constructive activity at two different scales. When a sequence is edited together out of a subset of the shared pool of video, the user is composing at a micro scale: she edits together her own sequence, and without contextual information, her sequence is isolated from those of other editors. If the interface could visualize the global collection of clips and sequences, along with contextual information about how the sequences were related to each other, the user could develop a sense of constructing at a macro level. Each new sequence would be shown not as an isolated piece, but as part of a larger communal creation. Communimage, described in the chapter 3, is an environment that supports both micro and macro construction: each user can create and submit a single digital image at a time (micro), but by placing new images next to existing images, each individual picture helps grow and evolve the larger mosaic (macro). The Individeo browser similarly attempts to represent constructive activity at both micro and macro levels.

Casual editing

The current video editing tools, surveyed in chapter 3, rely on the timeline or the grid as the principal interface for composing a sequence. One weakness of these interfaces is that it is not easy or efficient to quickly explore different ideas about how to edit a sequence. The user might want to add new clips or remove existing ones, at the beginning, end or middle of the sequence. She may also want to change the

order of clips in a sequence. Accomplishing these simple tasks can require very precise movement and repositioning of clips along a timeline or a grid: moving clips to a different position along the timeline will only be executed successfully if the mouse is precisely positioned while repositioning the clip. The user may also have to perform extraneous management of available space. For example, to swap the order of clips A and B within a sequence, the user may have to move clip B to a temporary location first, before clip A can be moved into clip B's old location. Individeo addresses these issues with a storyboard interface that supports what might be called casual editing, where the focus is on allowing editors to easily and efficiently try out various ideas for editing. Because the storyboard allows a sequence to be defined through a loose layout of clips within the editing space, there is no need for careful positioning. Being able to do this would be a strength in any editing interface, but in shared media construction, we are faced with a situation where the user is likely to be accessing unfamiliar clips coming from many other users. As such, it is deemed important to be able to quickly view and experiment with various options for how to compose a given sequence.

Communication with media

In most shared video systems, any type of communication among users is conducted in a text-based channel divorced from the video content. The user might choose to watch a particular clip or sequence; then, if she wants to comment on the viewed clip or send a message to the creator of the sequence, she must switch to a separate text-based medium such as message boards or chat rooms - if the application chose to supply such channels at all. In these text-based communications, there is no way to make direct references to the content being discussed. One must resort to phrases such as "the part in your sequence with the frontier psychiatrist saying..." or "the shot of her cancelling her flight tonight..." to refer to points of interest within the video content. Individeo attempts to provide a way of embedding commentary and other types of text-based communication directly into the video sequences. The user can talk about the content using the content itself, by composing messages in which text and video are mixed together.

4.2 Individeo browser

The Individeo browser occupies the lower-left portion of Individeo.

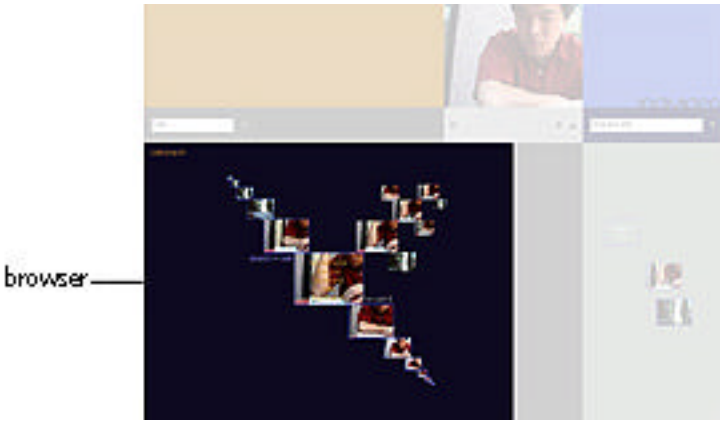
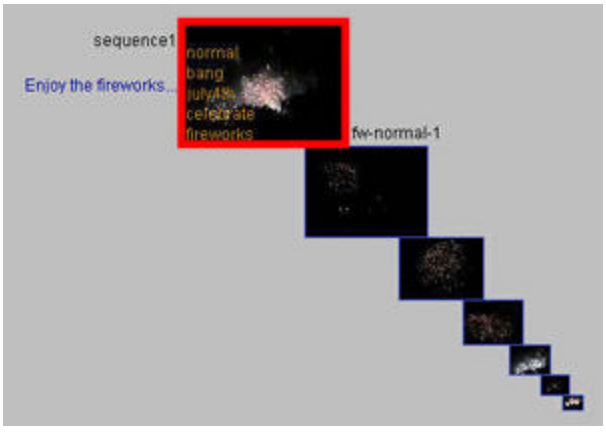
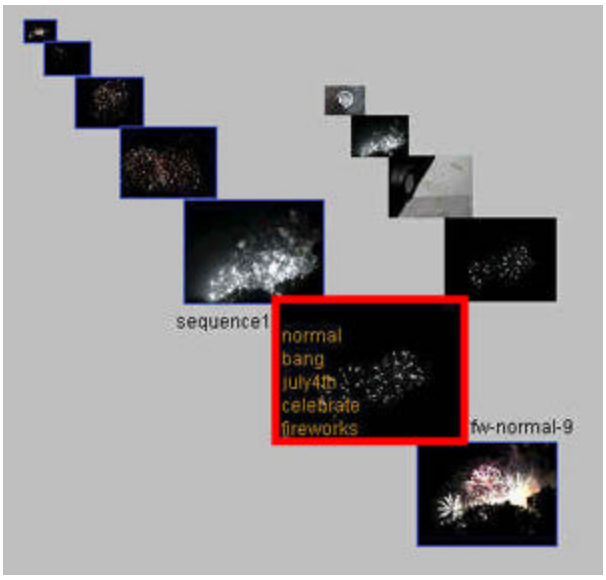


Figure 4-2. The Individeo browser.

Using the browser, the user can retrieve and watch clips and sequences, search for clips and sequences using keywords, and select clips to be edited with the Individeo editor. Inside the browser, each sequence is represented by the thumbnails of the clips in the sequence, arranged along the diagonal that runs from the top-left to the bottom-right. The first thumbnail from the top-left represents the first clip in the sequence. Every currently displayed sequence is represented by its own diagonally arranged set of clip thumbnails. Also, at any given time, there is a single focus clip within the browser, which is selected by the user.

a. 

b.



c.

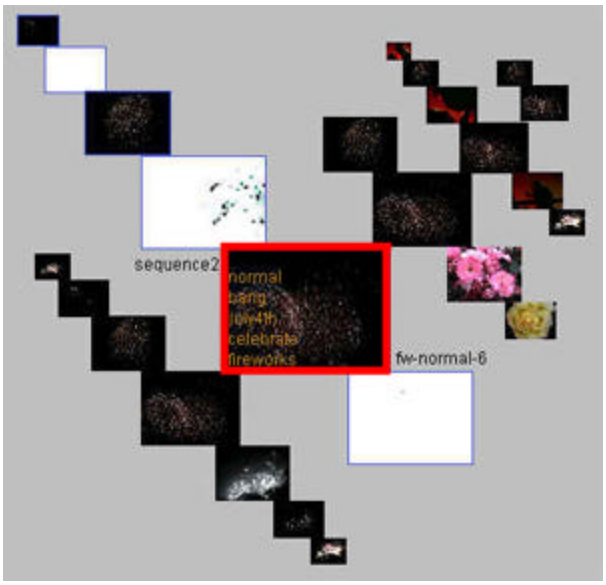


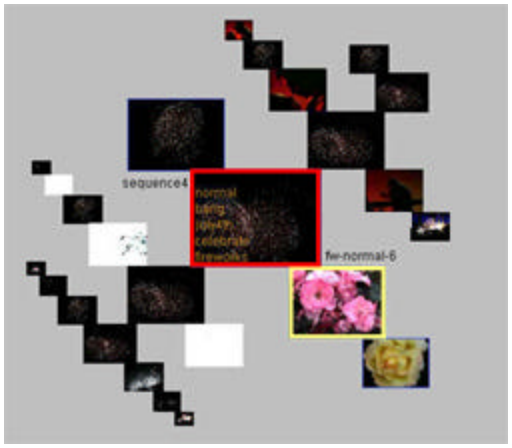
Figure 4-3. Visualizing video sequences in the Individeo browser.
The focus clips have been highlighted in red.
a. Displaying a sequence. The thumbnail at the extreme top-left position represents the first video clip in the sequence. Because it is the current focus clip, it is shown as the largest thumbnail.
b. Displaying two sequences. The focus clip is the next-to-last clip in the currently selected sequence. The focus clip is also part of one other sequence, and this other sequence is shown alongside the currently selected sequence.
c. Displaying five sequences. The focus clip is part of the currently selected sequence, plus four additional sequences.

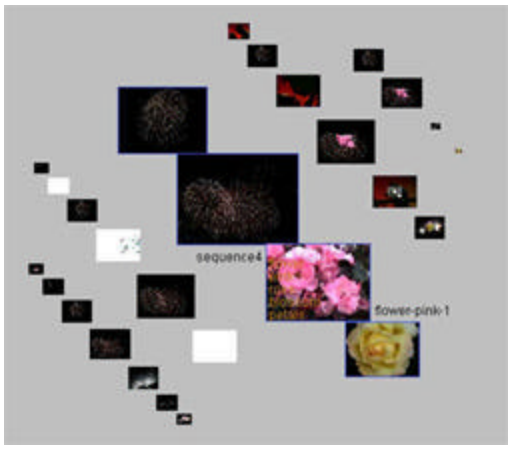
Visualizing context

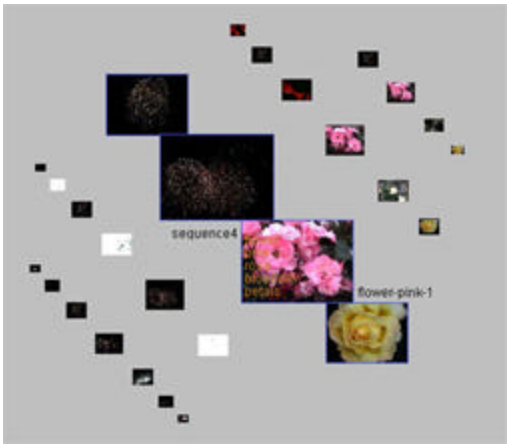
One of the primary goals of the Individeo browser is to visualize how multiple editors are reusing the same shared video content. It achieves this by displaying all the different sequences that contain a given video clip, so that the user can see how that clip has been reused. In the browser, the user selects a sequence by moving the mouse on top of any of the clips in the sequence. The selected clip becomes the focus clip, the current focal point of interest. The selected sequence is highlighted and animated towards the middle of the browsing area, and the focus clip is positioned in the center. Additional information about the focus clip is displayed around the thumbnail: the title of the focus clip title and the title of the current sequence. When the focus clip changes, the browser extracts information about all other sequences that make use of the focus clip, and displays those sequences. Each of the additional sequences are arranged to the left and the right of the currently selected sequence, along their own diagonals. The following figure illustrates the dynamic change of the browser depending on the user's selection of focus clips.

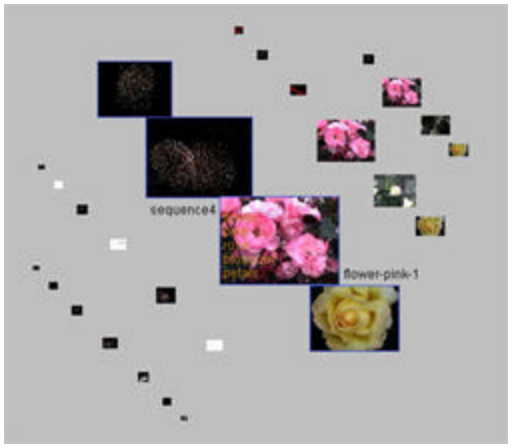
Figure 4-4. Interacting with the browser by selecting a focus clip.
The figures below (a - g) show step-by-step the animated transformation of the browser
when the user selects a new focus clip.

a. The current focus clip of fireworks is at the center of the browser, framed in red.
When the user selects a new focus clip of flowers, which is the next clip in the current sequence, framed in yellow,
the browser is updated through an animated visual transformation:

a. 

b. 

c. 

d. 

e. 

g. 

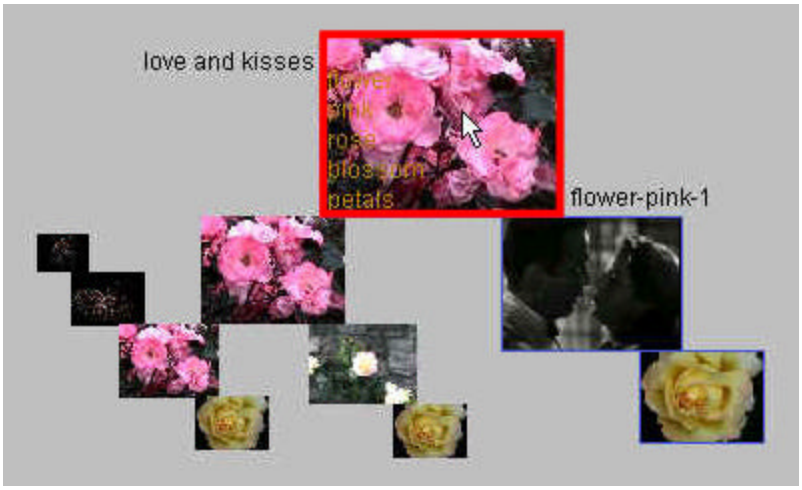
g. The animation has completed.
The new focus clip of flowers is now at the center of the browser, shown framed in red.

Each time the user selects a new focus clip, the browser updates itself, using a smooth, animated transition shown in the previous figure. The newly selected sequence is moved to the middle of the screen, and all other clips are resized properly to reflect the change in the focus clip. All clips that are no longer relevant (i.e. all currently visible clips that are not part of any sequences that contain the new focus clip) decrease in size gradually until they no longer appear onscreen. The animated transition is designed to prevent confusion about the change in the state of the browser: when the user interacts with the dynamic structure within the browser, and causes the browser to change by selecting a new focus clip, the animated transitions help maintain the context across different states of the browser.

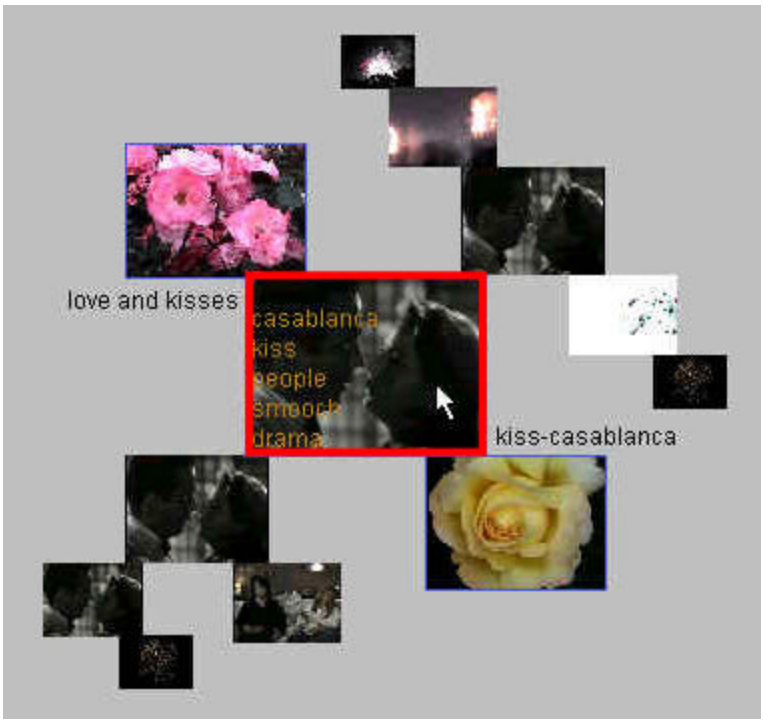
Once the browser has returned to a stable state, all sequences shown are connected to each other by their instances of the focus clip. As a result, the browser shows all instances of the focus clip along the top-right / bottom-left diagonal, running through the center of the browsing area. The following figures will illustrate a typical interaction with the browser, as the user navigates from clip to clip and from sequence to sequence. Each change in the focus clip causes an animated transition as shown in the previous figure.

Figure 4-5. Interacting with the browser.
The focus clips have been highlighted in red.

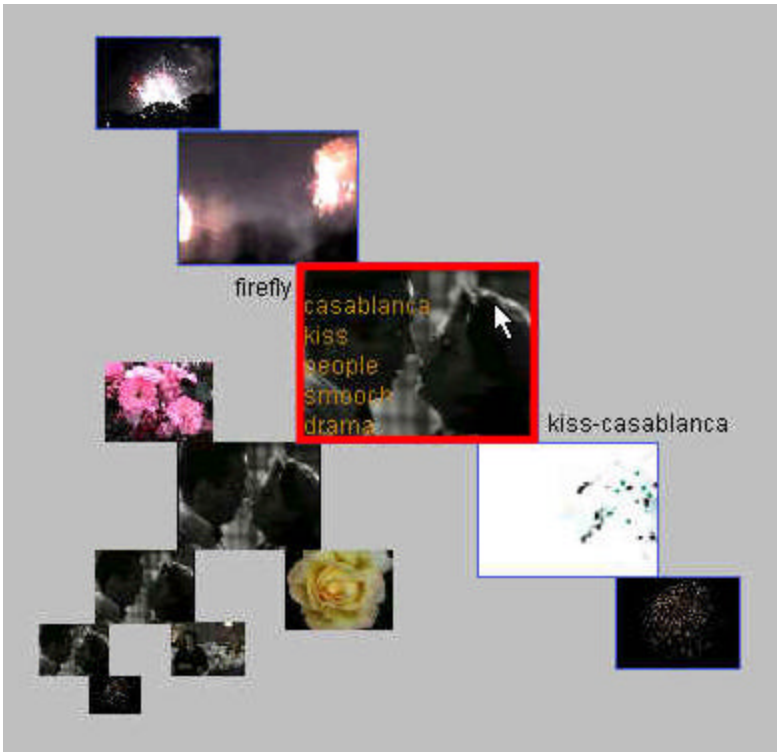
a.



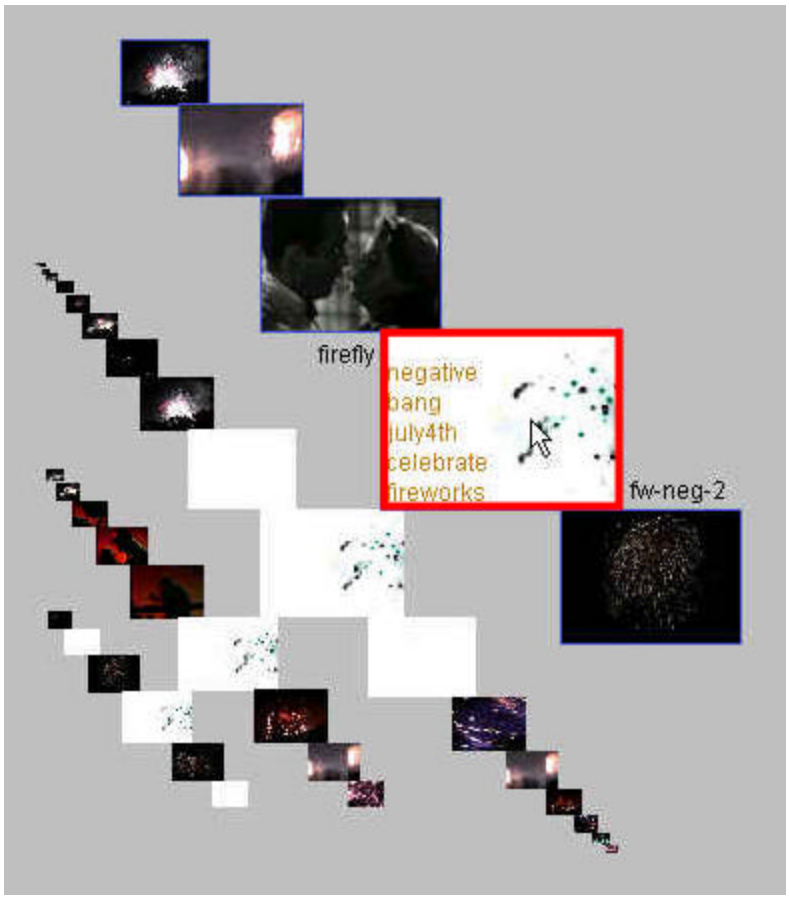
a. With the mouse, the user has selected "flower-pink-1" as the current focus clip.
The current selected sequence is "love and kisses".
The browser displays all the other sequences that also contain "flower-pink-1",
in this case two sequences,
shown to the left of the currently selected sequence.

b. 

b. When the user moves the mouse to the next clip in the sequence "love and kisses", it causes the browser to be updated with respect to the new focus clip.
The new focus clip is "kiss-casablanca", which is contained in three other sequences.

c. 

c. The user has navigated to a different sequence, "firefly", that also contains the current focus clip "kiss-casablanca". Although the current selected sequence has changed, the focus clip remained the same.

d. 

d. The user has moved to the next clip in the sequence "firefly", called "fw-neg-2".
Again the browser display is modified to show all other sequences that contain the new focus clip.

The browser uses a dynamic, animated visual structure to display all the sequences that contain the focus clip. A variant of generalized fisheye views [Furnas 1986] is used to arrange the clips and sequences within the browser. Using fisheye views for information display is an effective solution for two tasks: displaying a large amount of data within limited screen space, and visualizing the context of a particular data point with respect to the overall body of information. Both of these strengths were directly applicable to the goal of the Individeo browser, which was to display the possibly large number of clips and sequences inside the fixed space of the browser, in such a way that the user can derive contextual information about those clips and sequences.

Within the dynamic layout of the Individeo browser, the current focus clip is the largest clip thumbnail shown. Other clips in the current sequence decrease in size according to their position in the sequence relative to the focus clip. All other clips are resized accordingly, growing smaller in size the further they are

from the focus clip. The calculation of the clip sizes follows a modified version of the Degree of Interest (DOI) strategy as outlined for generalized fisheye views. Clips with higher DOI values appear larger in the browser display. The DOI values are calculated using the following steps in order:

Figure 4-6. Calculating the layout of clips within the browser. The figure shows the order of calculations of the size and position of each clip, as described below.

1. The highest DOI value is assigned to the focus clip, which is selected by the user as the current point of greatest interest.

2. The clips in the same selected sequence as the focus clip are assigned DOI values relative to the DOI value of the focus clip, depending on their relative position away from the focus clip. If a clip is further away from the focus clip within the sequence, it is assigned a lower DOI value.

3. The instances of the focus clip found in all other sequences are assigned DOI values also relative to the DOI value of the focus clip. The values depend on the chronological separation of respective sequences from the currently selected sequence. If a sequence was created much later than the currently selected

sequence, that sequence's instance of the focus clip will have a lower DOI value. The DOI values for these instances of the focus clip are weighed less than the clips in the same sequence as the focus clip.

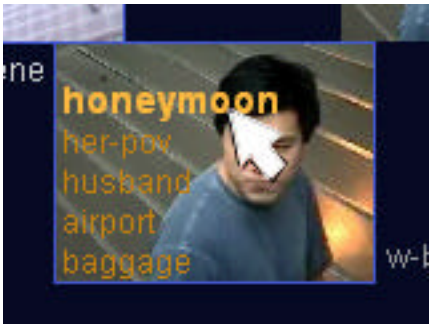
4. The other clips in sequences that contain the focus clip are assigned DOI values relative to the DOI value of their corresponding instance of the focus clip. This is similar to the DOI assignment of other clips in the currently selected sequence.

The Individeo browser builds on the earlier projects based on fisheye views, in order to supply both local and global information about how shared media is being reused. The browser provides local detail about the current focus clip and the selected sequence, and global context about the relationship of the current clip and sequence to other clips and sequences. The Individeo browser is responsive to real-time user input, similar to the Hyperbolic Browser (see chapter 3). The data for the Individeo browser is not a pre-defined hierarchy as in the Hyperbolic Browser; rather, a different set of information hierarchy is constructed and displayed when the user shifts the focus of the browser to another clip. Most significantly, because of the domain of video editing, the Individeo browser has a strong, explicit representation of the linearity of the information specific to this domain - the video sequences. The interface makes clear the precise subset of clips in a given sequence and the order of those clips, by arranging the clips belonging to each sequence along a unique diagonal line.

The browser allows the user to gain a broad understanding of how shared clips have been reused in various sequences. At a glance, they can review the other clips in sequences that contain the focus clip, and discover the different contexts in which the focus clip has been placed. The user can move from sequence to sequence by way of the connection of the shared focus clip. The activity of browsing becomes a constantly revealing experience, involving the discovery of relationships among sequences created by different editors. The global collection of sequences forms a visual collage on its own, because every time a new sequence is created, the browser updates itself to incorporate the new sequence. The next time one of the clips in the newly added sequence is selected as the focus clip, the new sequence will be displayed within the browser. Each sequence is therefore presented not as an isolated creation, but as part of the web of interconnected sequences. The user has a sense of adding to a larger whole at a macro level, every time she edits an individual sequence at the micro level.

Browsing and searching

The user can also search within the browser for video clips that share a common keyword annotation. The keywords are associated with the clip when the creator of the clip submits the clip to the shared video database. When the user selects a focus clip in the browser, its keywords are displayed inside the clip. The user can select one of these keywords to initiate a search for all clips that have been annotated with the same keyword. She can also directly enter the desired keyword into the search keyword field, located in the upper left region of the browser. Initiating the search through either of these 2 methods will cause a search to be executed within the shared video database.

a. 

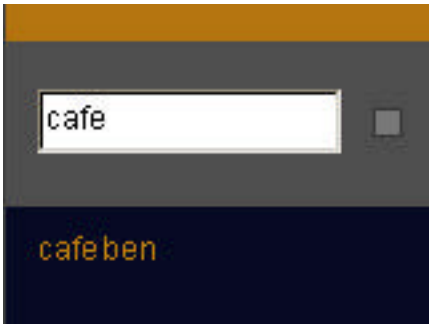
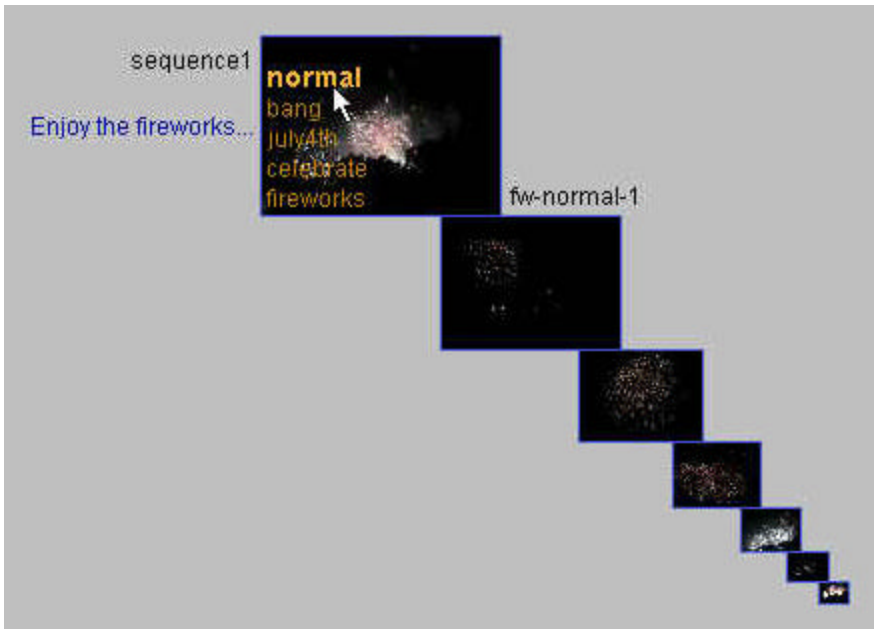
b. 

Figure 4-7. Two ways of initiating a keyword search.
a. Clicking on a keyword associated with a clip. Keywords are displayed inside the focus clip.
b. Entering the keyword directly into the search keyword field.

When a keyword search is performed, the results are displayed in the browser in the same visual form as a user-edited sequence. The results are shown as a diagonal line of thumbnails representing all clips that match the search keyword, with the first matching clip as the new focus clip.

a.



b.

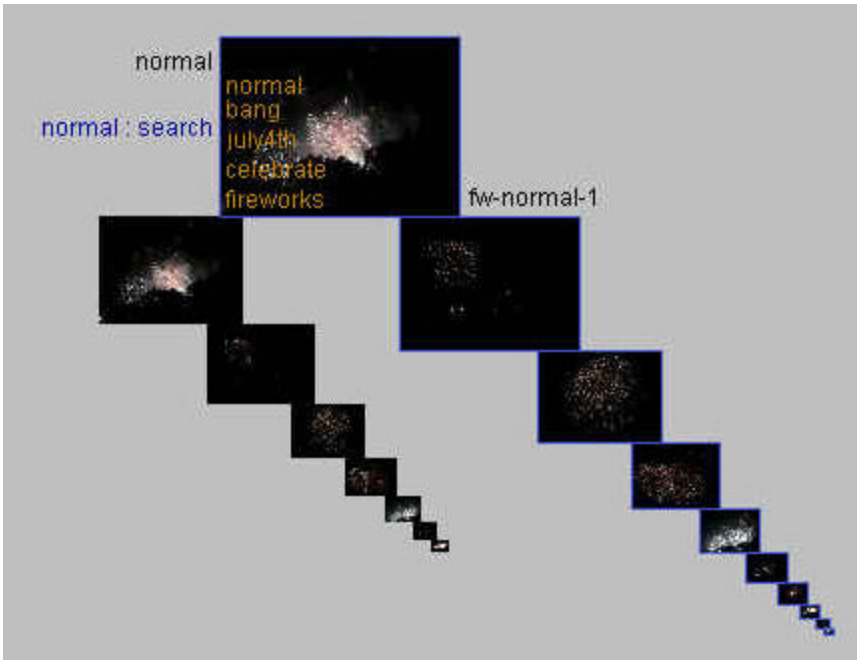


Figure 4-8. Before and after searching for clips using a keyword.
a. Before: the user initiates a keyword search, by selecting the keyword "normal" which is highlighted.
b. After: the search results for "normal" are returned in the same visual form as user-created sequences. The search result can be browsed and navigated like an ordinary sequence.

This provides increased flexibility in switching fluidly between the two modes of browsing and searching. In most applications that have some type of search functionality, the user must switch to a different interface to review the search results, typically a scrolling list of match results. With the Individeo browser, the user can perform a search and then continue to browse without switching between different interfaces. Put

another way, this mechanism can be seen as a form of query-initiated-navigation: each search returns structured information that can be readily navigated and browsed immediately.

A history of previous search keywords is displayed below the search keyword field, and selecting any of these shortcuts will retrieve a previous search result.

4.3 Individeo editor

The Individeo editor occupies the lower-right portion of Individeo.

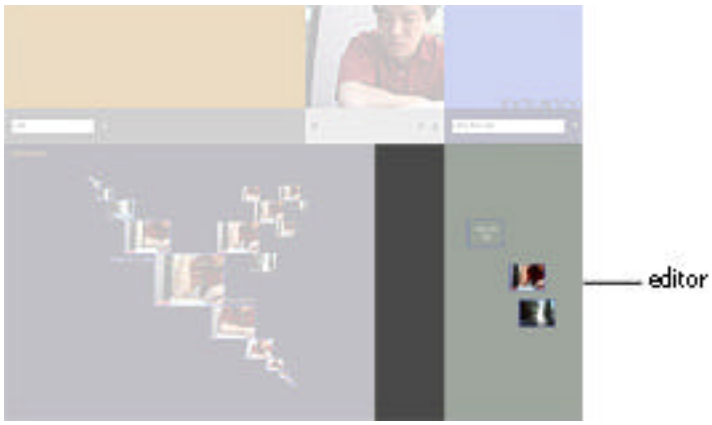


Figure 4-9. The Individeo editor.

Once the desired clips have been selected from the browser, the editor allows the user to specify the order of clips within a sequence, add text clips to the sequence, preview the sequence, and upload the sequence to the shared video database.

Storyboard editing

The editor features a storyboard interface where the user can try out different ideas for how to edit a sequence, in terms of which clips to include and what order they should be played in. In film and animation

production, pictures representing a shot or a scene are laid out on the storyboard. While formulating how a segment could be edited, the editor adds or removes pictures, and rearranges them to understand the order of shots and scenes. Pictures can be set aside temporarily while considering an alternate edit, or can be overlapped with each other to indicate sub-units within the overall segment. The Individeo editor is an attempt to capture the casual feel of brainstorming with a storyboard.

To begin editing a sequence, the user can select and drag over from the browser the clips they want to edit with. Once the selected clip is dropped within the editor, it is added to the editor interface as an editor clip, shown as the thumbnail image of the clip. The editor clip can be freely repositioned within the editor. Moving the mouse on top of an editor clip displays the clip's title. Within the editor, there are 2 regions, differentiated by their color; the lighter-colored region is the sequence space. Only the clips that are placed within the sequence space are considered as part of the sequence being edited. Clips can be left outside the sequence space, but still within the editor. By moving clips in and out of the sequence space, the user can choose to temporarily add or remove a clip from the edited sequence, while she considers how the sequence should be edited.

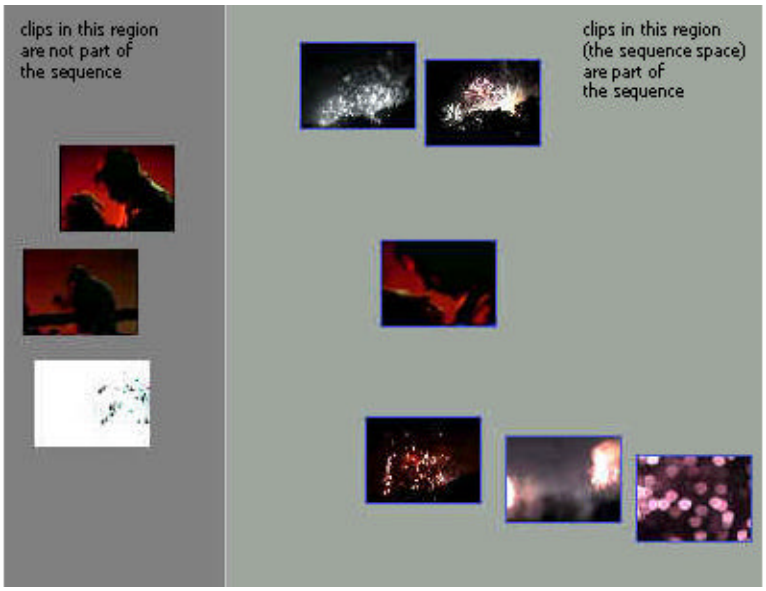


Figure 4-10. The editing area, split into two sections.
The lighter-colored region on the right is the sequence space, and only the clips within this region are considered part of the sequence being edited. Clips can be set aside outside the sequence space while different editing options are considered.

Editing a sequence means simply arranging the clips in the desired order within the sequence space. Unlike in a timeline with only one horizontal dimension for arranging the video clips, the Individeo editor clips can be arranged spatially in two dimensions. The order of the clips in the sequence is automatically computed by the editor. The editor will parse the layout of the clips, following the traditional Western rules for reading a comic book, from top-left to bottom-right. The clip in the furthest top-left position becomes the first clip in the sequence, followed by the next clip to the right and to the bottom, and moving on down to the clip in the furthest bottom-right position. Horizontal ordering takes precedence over vertical ordering.

Figure 4-11. Editing a sequence by arranging the clips within the sequence space.
The numbers have been added to show the ordering of the clips in the sequence.

Visual layout of editor clips defines the sequence, and changing the precise set of clips in the sequence or the ordering of those clips is simply a matter of rearranging the layout. The editor does not require the user to be precise with the positions of the clips, and they can be even be overlapped. One benefit of this lack of constraint is that the visual layout can be used to organize the clips in a sequence in any way the user sees fit. For example, the following two sequence-space layouts define the same sequence. The first layout might be the result of a quick brainstorming session, with clips loosely spaced out to clearly show their order. The second layout shows a more careful clustering of clips into five separate horizontal groupings, each group

representing a subsection of the sequence. The user can visually organize her shots within the sequence space as she sees fit.

a. b.

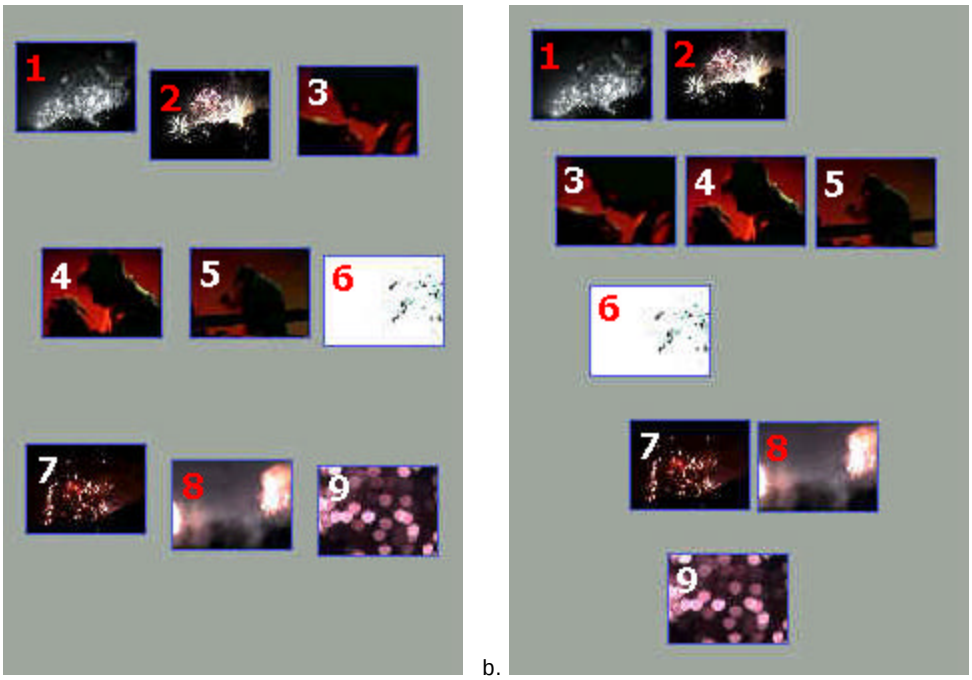


Figure 4-12. 2 different layouts representing the same sequence.
The numbers have been added to show the identical ordering of the clips in both layouts as determined by the editor's parsing of the visual arrangement of the editor clips.

Video and text

A sequence in Individeo can contain text clips in addition to video clips. To create a text clip, the desired text is entered into the text clip field in the upper right region of the editor. When a text clip is created, it is added to the sequence space in the form of a clip rectangle containing the text. Text clips can be moved and placed inside or outside the sequence space, just like video clips.

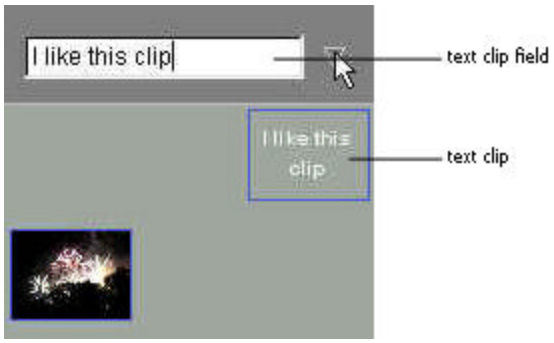


Figure 4-13. Adding a text clip in the Individeo editor.
When the text is entered in the text clip field, a text clip is created.

When a sequence is played out, text clips play back in the form of inter-titles, in white text against black background. These inter-titles are similar to ones found in classic silent films. The text inter-title format helps juxtapose video and text within a sequence. Inter-titles are flexible enough to accommodate a wide range of uses, such as explanations of editing choices in a sequence or pointers to upcoming highlights within a sequence. In discussions about a sequence, the comments are no longer separate from the video, but embedded in the video sequence itself.

4.4 Individeo player

The video player is shared by both the browser and the editor. There are play and stop buttons for the player on both sides of the player window, so that it can be used with both the browser and the editor. For browser-side playback , the currently selected sequence containing the focus clip is played back; for editor-side playback, the current sequence as defined by the clips in the sequence space is played back.

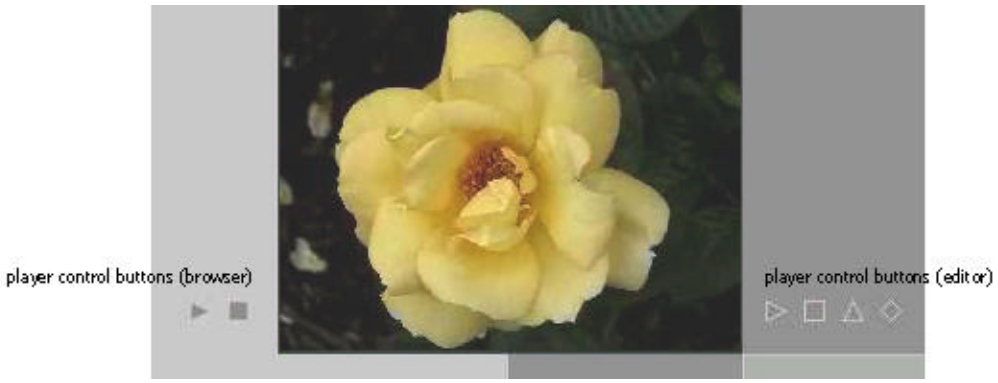


Figure 4-14. Individeo player, used to play back sequences from both the browser and the editor. There are controls on either side of the player, so that it can be used to play back sequences from the browser or the editor.

4.5 Shareable Media

Individeo is part of the Shareable Media project, a joint effort by several graduate students in the Interactive Cinema group. The goal of Shareable Media is to build an infrastructure for supporting online applications for rich media such as video and audio. By using the components of the Shareable Media project, collaborative video editing tools like Individeo can provide custom interfaces on the user end, while relying on a standardized API for back-end content management. Shareable Media is composed of 3 tiers: Data Tier for low-level databases that store video content and user information; Business Logic Tier for application interfaces and manager components; and Presentation Tier for user-defined front-end interfaces such as Individeo. The Business Logic Tier contains the Shareable Media Framework, a Java-based API which application developers can use to communicate with the back-end servers. Within the same tier, the Application Manager component handles the registration and configuration of front-end applications like Individeo.

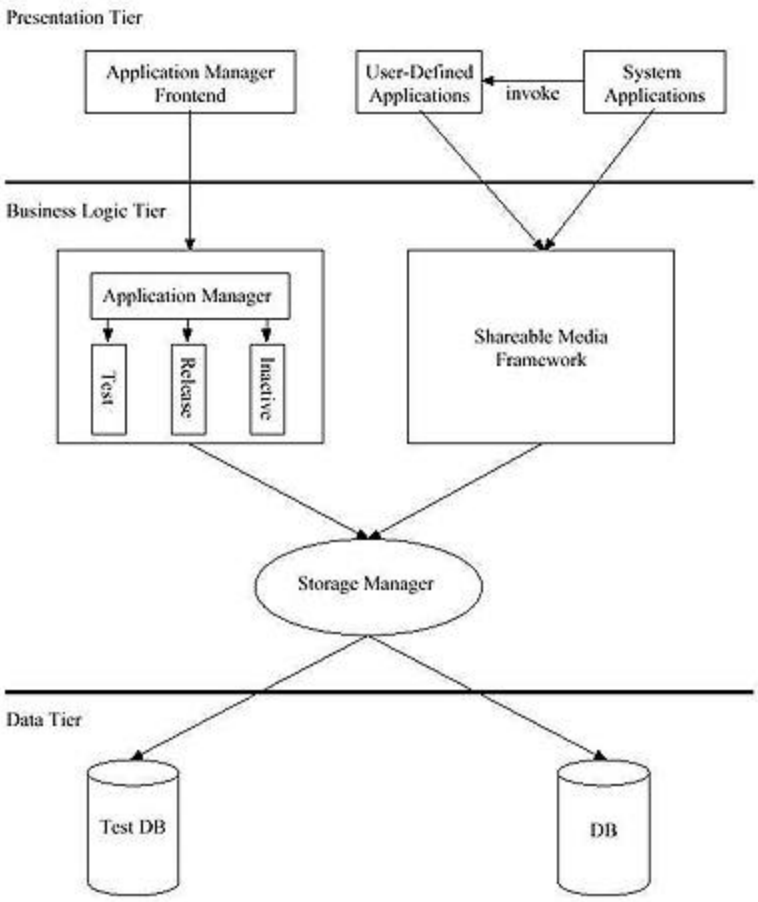


Figure 4-15. The 3-tier Shareable Media Architecture.

Individeo is implemented as a Java applet. The Individeo applet is fully compliant with the Java Development Kit (JDK) 1.1 standard from Sun. Since JDK 1.1 is widely supported today by leading browsers such as Microsoft Internet Explorer and Netscape Navigator, online users can run Individeo without the use of any custom plug-ins.

Individeo relies on Apple's QuickTime technology for controlling the playback of video clips. To play back a sequence of QuickTime video clips one after another, Individeo uses the Synchronized Multimedia Integration Language (SMIL), a Web-standard markup language for multimedia, recognized by the World Wide Web Consortium. A SMIL document contains the specifications for playback of a sequence of media elements including video, audio and text. When the user asks for a sequence to be played back, the Individeo applet sends a request to the Individeo servlet residing on the Shareable Media server. The servlet dynamically generates and returns a SMIL file which contains the specifications for the requested sequence.

The returned SMIL file is then loaded by the QuickTime plug-in, and the sequence of clips is played back. Since SMIL supports a variety of media types, it supports the Individeo sequences containing both video and text.

4.6 Summary

This chapter presented the design issues and interface components of Individeo, a tool for online sharing and editing of digital video. Here I summarize the four design issues outlined at the beginning of the chapter, and briefly describe how Individeo's interfaces address these issues.

Shared contexts

Users who are sharing video clips may edit a given video clip in different ways, each edit putting the clip in a new context. The Individeo browser attempts to visualize these contexts and provide the editors with contextual information about their clips and sequences. The browser can simultaneously display all sequences that contain a particular clip, and therefore can show how sequences are related to each other by their commonly used clips.

Scales of construction

When a user composes a sequence with shared video, she is engaging in two levels of media construction: she is constructing her own sequence independently (micro level), and she is also adding to the global sum of all sequences that use the shared media (macro level). In the Individeo browser, both levels are shown in one unified interface: a sequence is clearly represented by the clips that make up the sequence, and it is always displayed alongside other sequences that share common video clips. The creative output of each person is shown as linked with those of other users.

Casual editing

While editing, the user may wish to brainstorm with different ideas about how to edit a particular sequence: she may try out add, remove and reorder clips in a sequence in quick succession. The Individeo editor allows the user to compose a sequence by loosely rearranging the clips inside an editing region, without requiring precise positioning of clips when defining the sequence. Various ideas on how to compose a sequence can be tried out quickly.

Communication with media

While collaborating on a project, users may want to comment on each other's sequences or discuss their ideas on editing a sequence. The Individeo editor offers the capability of embedding text within a video sequence in the form of inter-titles. This simple device allows collaborating editors to create sequences that mix video and text, and serve as mixed-media messages where the text can serve a variety of purposes. For example, the text could refer to and highlight a certain portion of the sequence.

5. Evaluation: Honeymoon

See beings and things in their separate parts. Render them independent in order to give them a new dependence. | *Robert Bresson, Notes on the Cinematographer*

Honeymoon is an experimental digital video production in which two geographically separated videomakers collaborated online to tell a story about a newlywed couple on their honeymoon. Two teams of directors and actors shot footage in two different cities, and the captured footage was then shared and edited online with Individeo, the collaborative video browsing and editing tool described in the previous chapter. The production resulted in a series of video sequences that presented disjointed fragments of the narrative, which can be explored using Individeo. Honeymoon served as an evaluation of how Individeo's custom interfaces can affect the process of close collaboration by filmmakers.

5.1 Why Honeymoon?

Honeymoon was conceived as a very specific and constrained evaluation of Individeo. There are many different application scenarios in which users can edit and share media with Individeo. These scenarios involve a variety of types of collaborations: a fixed number of partners might be working closely on a well-defined project, or a very large community of editors may be discovering and reusing each other's video over a long period of time. Honeymoon is an attempt to explore one particular type of collaboration, between two groups who are shooting and editing almost simultaneously, having agreed upon on the story outline. Although this is a particular type of collaboration, using Honeymoon to evaluate Individeo would lead to issues and observations pertinent to a wider range of collaborative activities.

I chose Honeymoon out of my desire to use a tool I developed towards telling a story with my own content. John Maeda writes in his reflections on building an animation system [Maeda 2000]:

> The dilemma for a person who builds tools for others is that she rarely learns how to use it for herself. As a selfless virtue, creating new tools for others is a commendable public service that has its own set of rewards, sometimes monetary, in which case the virtue value lowers proportionately. One can argue that the design and implementation of a software tool is by itself an extremely creative, artistic activity. History shows, however, that it is the artist who first effectively uses a new tool, not the person that makes the new tool, who is remembered.

Individeo was originally designed as a "tool for others": its goal was to support the sharing and editing of video within a large community of users. With Honeymoon, I saw an opportunity to try using Individeo as a tool for myself, for telling a story I felt passionate about. By engaging in producing Honeymoon with Individeo, I wanted to evaluate my software as an active user, and discover how to use the tool for myself.

5.2 Story outline

Honeymoon is a story about a newlywed Asian-American couple, who are on their honeymoon in New York City. On the surface, they seem like every other newlywed couple. Slowly, however, a distance grows between them. Its roots could lie in a well-kept secret, best left unexposed; in a word and a gesture, the smallest of signs; or the discovery of incompatible dreams about their new life together. Isolated in their strange surroundings, they only have each other, but a rift begins to develop between them. Honeymoon is a story about this rift between two people.

The couple has brought a video camera for the trip. They use the camera like any tourist might, shooting video of the sights and of each other. As they slowly grow apart, the husband and the wife continue to shoot each other with the camera, but increasingly in unexpected ways. The camera is left on

surreptitiously, recording the subject without his or her awareness. The digital effects are used to stretch, discolor and otherwise distort the subject's image, as if venting anger towards the subject through the image. Cheerful messages on street signs are framed in the same shot as the moody visage of the person being captured.

In these and other ways, the husband and the wife begin to rely on the video camera as a device to document their honeymoon from each person's point of view, and to express what cannot be discussed out loud. An alternate record is born, different from the "official" honeymoon video of a happy couple. The familiar device of the camera becomes a private tool for mediating thoughts and emotions that cannot be expressed otherwise.

5.3 Production framework

The story for Honeymoon came to me while I was developing Individeo. Afterwards, as I was searching for a suitable evaluation project for the Individeo interfaces, I realized that Honeymoon could make a good candidate. I contacted Phillip Tiongson, a New York-based filmmaker, Media Lab alumni, and close friend; he became my partner in the project.

The framework for the collaborative production of Honeymoon was defined as follows:

- The story is told strictly through the optical point of view of the husband and the wife, as he or she is holding the video camera. In other words, the footage that Phillip and I shoot will always take the perspective of the offscreen character, who is operating the video camera and shooting the video of the onscreen character. There may also be footage shot from the point of view of the camera alone, if the camera is recording on its own. There would be no footage from a third-person point of view. I was interested in this constraint, because I wanted to show how each of the two characters begins to (mis)use the video camera towards their own ends. The restriction would force me to present the story

solely through how the camera is used by the characters. I am a firm believer in the notion that movies should show rather than tell, and I saw this constraint as leading to a direct and immediate mode of cinematic storytelling.

- The production would be conducted by two teams who are each in a different location. Phillip would shoot in New York, where he lives and where the story is set in; I would be in Boston, where I live and work.

    - Phillip would direct and shoot in New York with an actress who portrays the wife. Since the story is told strictly through the point of view of the offscreen character, Phillip would effectively be "playing" the husband's role, as he shoots the video of the wife.

    - I would direct and shoot in Boston with the actor who portrays the husband character. Since the story is told strictly through the point of view of the offscreen character, I would effectively be "playing" the wife's role, as I shoot the video of the husband.

- After each shooting session, Phillip and I would each edit our footage into small clips, and share it with the other person using Individeo. We would then edit sequences in Individeo using our combined footage.

At first glance, such a framework might seem nearly unworkable. After all, the actors are never in the same place at the same time. For instance, this means we could not shoot a conversation between the two characters. Even if the director "plays" the role of the offscreen character in terms of operating the camera, he obviously cannot speak aloud in the role of that character. However, basic cinematic techniques can be relied upon to edit a conversation scene. Even in a traditional film production, a conversation between two characters may not be shot in a single take where the two actors are actually talking to each other. The setup known as shot / reverse shot will show one character alone in the frame, followed by the other character alone in the frame, then back to the first character, and so on. If performed and edited properly, the back-and-forth cutting between the two characters creates the cinematic illusion of shared time, space and conversation, even if the actors were shot in different takes - or in different places. Phillip and I could

shoot our respective actors accordingly, so that the characters seem to be talking to each other when the combined footage is edited into one sequence.

Reactive production was a major component of this collaboration, where Phillip and I would plan our shoots by reacting to the most recent footage from the other person. At the very beginning, we would try to shoot clips that allow us to make the sequences we had in mind, and to supply clips the other person may be able to use. Such clips might include blank gazes or close-ups, clips that are ambiguous enough to be used in a wide variety of sequences. Then, after successive shooting sessions, we would share our new footage with each other. The person who received new footage would review the new clips, and decide on how to react to that footage. For example, I could pick up on a particular action that the wife performs in one of Phillip's clips, and shoot footage of the husband commenting on that action. It became natural to take turns shooting, and building on the most recent shoot as we went along. With each turn, we would look for ways to support or contradict the story's direction.

It should be noted that the end result of Honeymoon was not planned to be a linear film, in the form of a two-hour feature or a fifteen-minute short. In Honeymoon, I was less interested in sequences that looked and sounded like the result of a typical production. Given that we were engaging in a long-distance collaboration using an unorthodox process and custom tools, I wanted to edit and produce content that reflected the process and the tools. I also wanted to have Individeo be the tool for viewing the story as well as for editing it together. Therefore, I chose to tell the story through a number of short, disjointed sequences. Each sequences would stand on its own, revealing a character trait, indicating some emotion, or advancing a plot point. For the viewer, the sequences would not necessarily be watched in any particular order, but rather the sum of all the sequences would comprise the narrative of these two characters.

5.4 Production results

The Honeymoon production was conducted through several shooting sessions occurring between April and July of 2001. Phillip Tiongson was in New York with Tina Lee, who played the role of the wife; I was in Boston with Dong Shen, who played the role of the husband. A total of 312 clips were produced by the two videomakers. Although the production halted prematurely due to scheduling issues, and the story could not be explored fully, I was able to gather enough footage to edit sequences and perform a limited evaluation of the Individeo interfaces.

There were technical issues which prevented the use of the networked version of Individeo. Due to bandwidth issues and problems with the New York ISP company, Phillip and I were unable to use the Web-based applet version of Individeo. I was forced to use the local application version of Individeo, which as the same interfaces as the Web-based applet version. The only major difference was that we could not rely on the Shareable Media architecture for storing and managing the clips and sequences. Instead, we worked around the problem by using local files.

- For clips, we would transfer the clip files manually after each shoot, so that we would have the identical set of clips on our respective machines. A text file was used to store the list of all clips and the keywords for each clip. Individeo would open this file upon launching, in order to load all the clips and their keywords.

- For sequences, I added additional code so that sequences were stored in a text file, with a special format for specifying the precise list and order of clips in each sequence. When Individeo was launched, it would load the sequence information from the text file; when a new sequence was created in Individeo, it would save the sequence to the text file. By exchanging the local text file, Phillip and I had access to the complete and identical set of sequences.

The use of the local application version added additional overhead for Phillip and me in terms of maintaining identical sets of clips and sequences on our machines. This did not affect the evaluation in any significant

way, since the interfaces for both the Web-based and local versions of Individeo were identical, and the focus of the evaluation was on the use of the interfaces.

5.5 Production examples

Here I present specific examples how I used Individeo to manage, share and edit the Honeymoon footage produced by both Phillip and myself. General observations drawn from these and other instances are presented in the following section.

Editing a new sequence

I thought that one starting point for the story could be the husband teaching the wife how to use the video camera, and introducing her to the various features of the camera. I imagined that later, we would see the wife using these same effects to distort and otherwise play with the captured image of the husband. I shot several takes with Dong where he speaks directly into the camera and explains how to use its features. After the shoot, I edited several short clips, added them to the shared pool of clips, and defined appropriate keywords for each clip.

I launched Individeo, and searched for the keyword "camera" to retrieve all the clips I had just edited. Individeo displayed the search results in the form of a sequence, which I could then browse through. By playing the "sequence" composed of the search results, I could automatically review all the clips in one sweep.

Figure 5-1. Search result for "camera".
Please note that all figures have been scaled down to save space.

I selected several of the clips to include in a sequence where the husband teaches the wife how to use the camera. I dragged these clips over into the editor. I tried several different combinations and orderings of the clips, in order to create a satisfying sequence. After testing several options, I decided on one sequence of clips. In the sequence, the husband starts by telling the wife about the basic controls of the camera such as the stop / start button for recording. He advises her about keeping the camera still and not abusing the zoom, which she promptly does much to his amusement:

Figure 5-2. Clips for the "camera" sequence, arranged in the editor.

I added this sequence to the shared pool of sequences. The browser updated itself and displayed the new sequence.

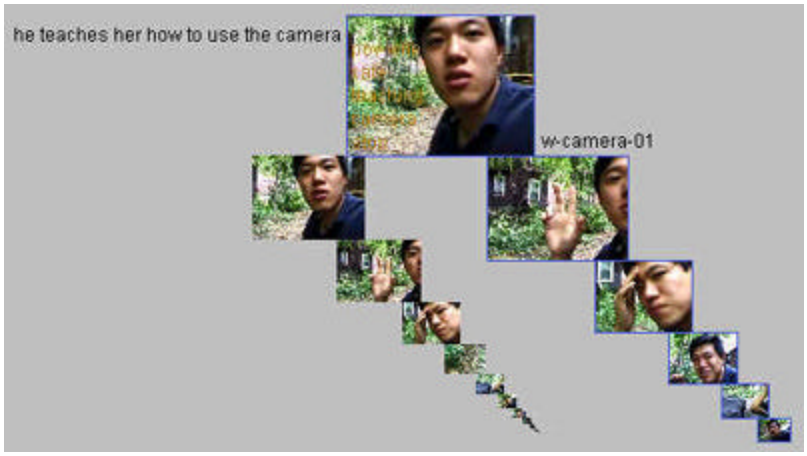
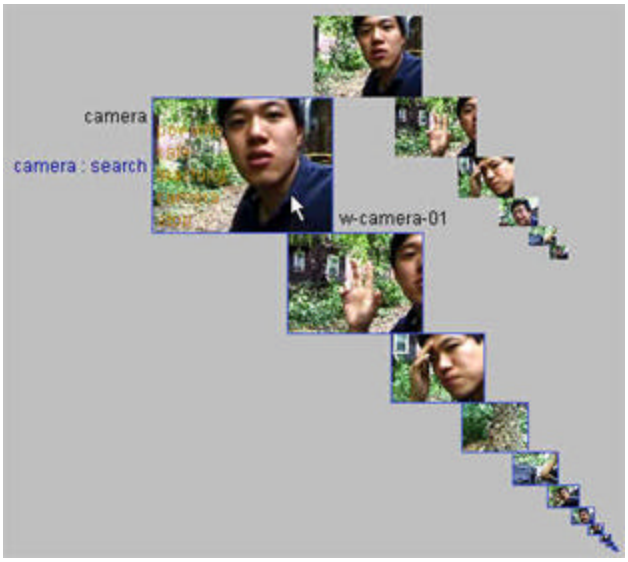
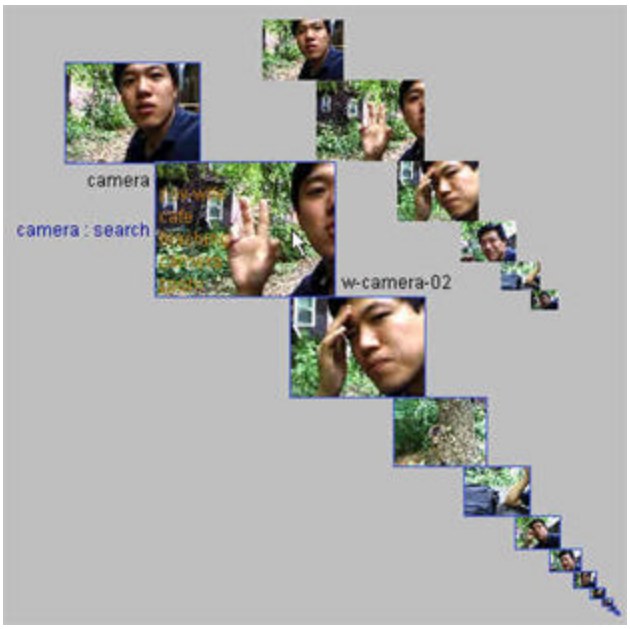


Figure 5-3. The "camera" sequence, shown in the browser.

Now when I browse through the earlier search results, my new sequence would be displayed whenever one of the clips I had used became the focus clip.

a. 

b.



c.

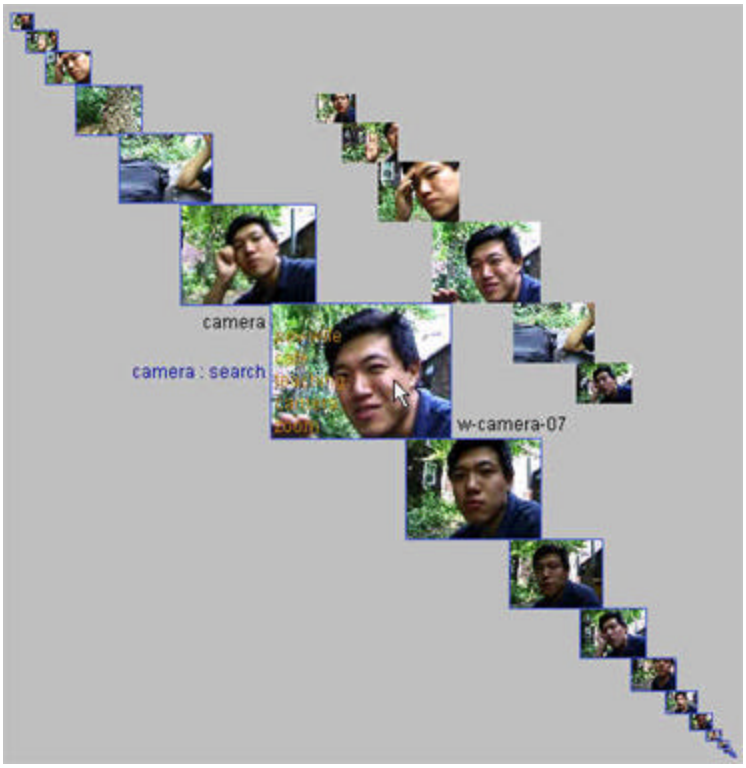


Figure 5-4. Three different views of the browser, where the user is browsing the search result for "camera". The newly added "camera" sequence is displayed alongside the search result whenever a clip in the sequence is selected as the focus clip.

Grouping similar clips into sequences

While editing, I realized that I wanted to keep track of certain clips together as a group. For example, in editing my sequence, I used a clip of the husband commenting about how to use the camera zoom. There were many other clips with the same content, since I had done several takes with my actor. Doing a search for "zoom" returned too many of these clips to review each time, and it was difficult to remember which ones I had watched and liked. I wanted to be able to create and maintain groups of similar clips, so that I could later access them when I needed that particular type of clip. I also wanted a way for Phillip to be able to see these groups of shots, so that he could understand what other shots I could have used to build a particular sequence. If he could quickly review all other clips where the husband talks about the zoom, he would be able to suggest alternatives to the clip I had chosen to use in my sequence.

I realized that I could create clip groups by simply creating sequences with the clips in question. I returned to my clips involving the camera. Using the same procedure I had used in making my earlier sequence, I created two additional sequences: one for clips where the husband talks about keeping the camera still, and one for clips where he talks about the zoom feature. I ordered the clips in the order of preference, so that the first clip in the sequence was the one I considered the best. These were therefore not typical sequences, but more like a mini-bin of similar clips I could access later. The browser was automatically updated to show my previously created sequence as well as my two clip-group sequences:

a. 
he teaches her how to use the camera
w-camera-02

b. 
clips: zooming in and out
w-camera-02

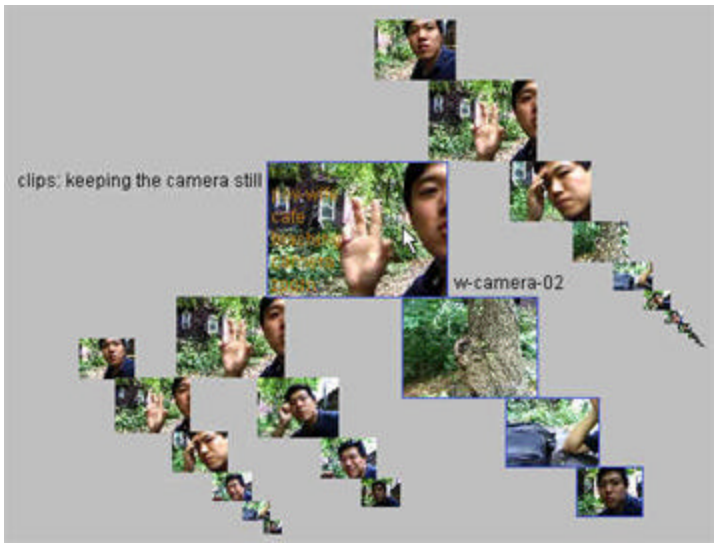
c. 

Figure 5-5. Three browser views of sequences of camera clips, each with a different selected sequence:
a. The "camera" sequence edited earlier.
b. Clip group for clips related to zooming.
c. Clip group for clips related to keeping the camera still.

Sharing and reusing clips

There were numerous instances where Phillip and I would use each other's clips to edit a sequence. The process of doing so was similar to how I would edit with my own new clips, as described previously. First, after I had received new clips from Phillip, I would search for particular keywords he had used. Then I would react to his latest footage, developing ideas for shots that I could combine with his clips in order to compose sequences. After shooting, I would edit the sequence and share it with Phillip.

In one instance, I searched for "women", retrieving Phillip's clips where the husband (whose point of view Phillip was shooting from) looks at other women while the couple is walking around the streets. This is sometimes noticed by the wife, much to her disapproval. Reacting to his footage, I shot some footage of the husband being distracted by other women on the streets, with the camera (now from the wife's point of view) catching him in the act.

a. b.

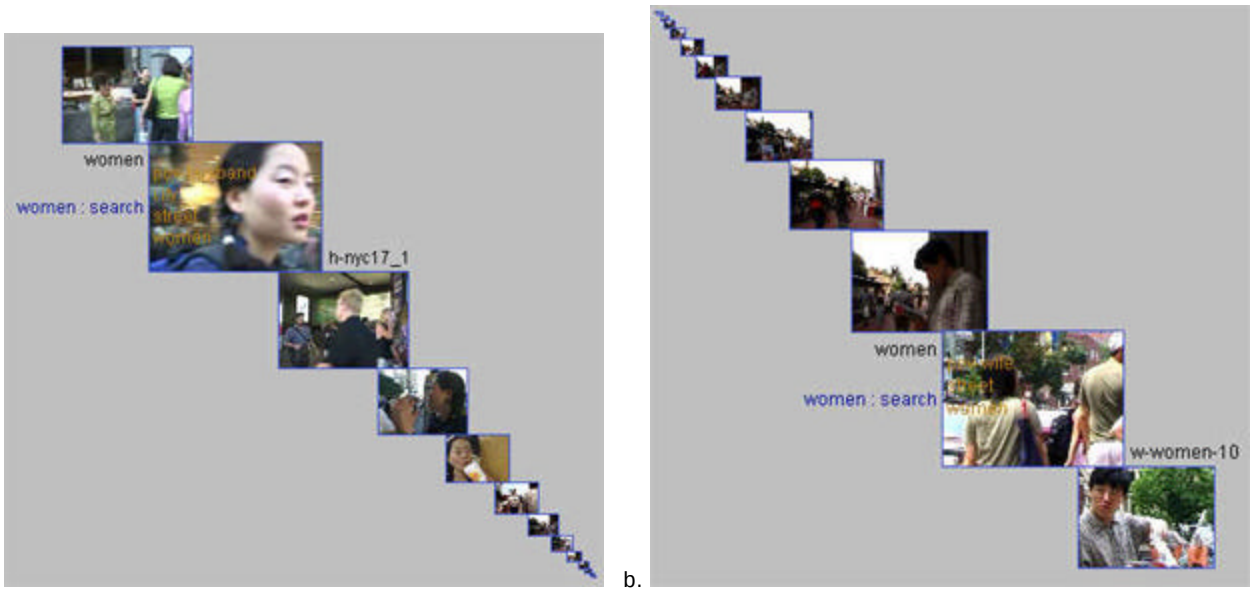


Figure 5-6. Two sets of clips about the husband looking at other women.
a. Phillip's clips.
b. My clips.

I combined these clips and edited a sequence to show the husband's distraction from both points of view.

Figure 5-7. The "women" sequence, shown in both the browser and the editor.

Similarly, Phillip also provided me with clips of the wife staring at mothers and young children. After looking at his footage, I wanted to explore how the couple might deal with the idea of having a child. In my next shooting session, I captured the husband talking to an off-camera person, supposedly sitting across the

table from the wife. The camera was positioned to make it appear as if it was recording the conversation without his knowledge, as if the wife had left it on secretly. Since the wife was obviously not present, Dong improvised his side of the conversation, in which he brings up the subject of having a child. I asked him to express several different attitudes and feelings towards the idea of having children, both positive and negative. I then edited together sequences representing these attitudes. In one sequence, the husband is pushy and aggressive about having a child as soon as possible; in another, he is concerned with her feelings, asking for her thoughts and reassuring her that he would be there for her no matter what.

Figure 5-8. The "baby" sequences. The selected sequence is titled "he calms her down about having a baby". The browser shows that all the sequences begin with the same clip.

Then, I created additional sequences by combining my clips of the husband's conversation with Phillip's clips of the wife looking at mothers and children. Phillip's clips were relatively neutral: the wife shows no explicit or exaggerated emotion upon seeing the families. Therefore, the wife clips could be suitably intercut with different types of the husband clips, whether he was being pushy or supportive. When the two sets of clips were combined together, the juxtaposition led to the wife clips seemingly reflecting her feelings about having a child, even though they were largely neutral.

Figure 5-9. The new "baby" sequence, titled "he talks about having a baby",
which contains clips from both Phillip and myself.
The browser shows that the current focus clip, w-baby-16, is used in two other sequences.

The next and last example of reusing each other's clips involves reactive production on both sides. First, I had shot some footage of the husband sitting at a cafe, looking at a New York City tour guidebook, and considering various locations for sightseeing. He is supposedly talking to the wife, describing the landmarks listed in the guidebook and considering the various walking paths. In reaction to the cafe clips, Phillip later shot and sent me several shots where the couple are exploring New York City on foot. The camera captures the wife as they are roaming the streets.

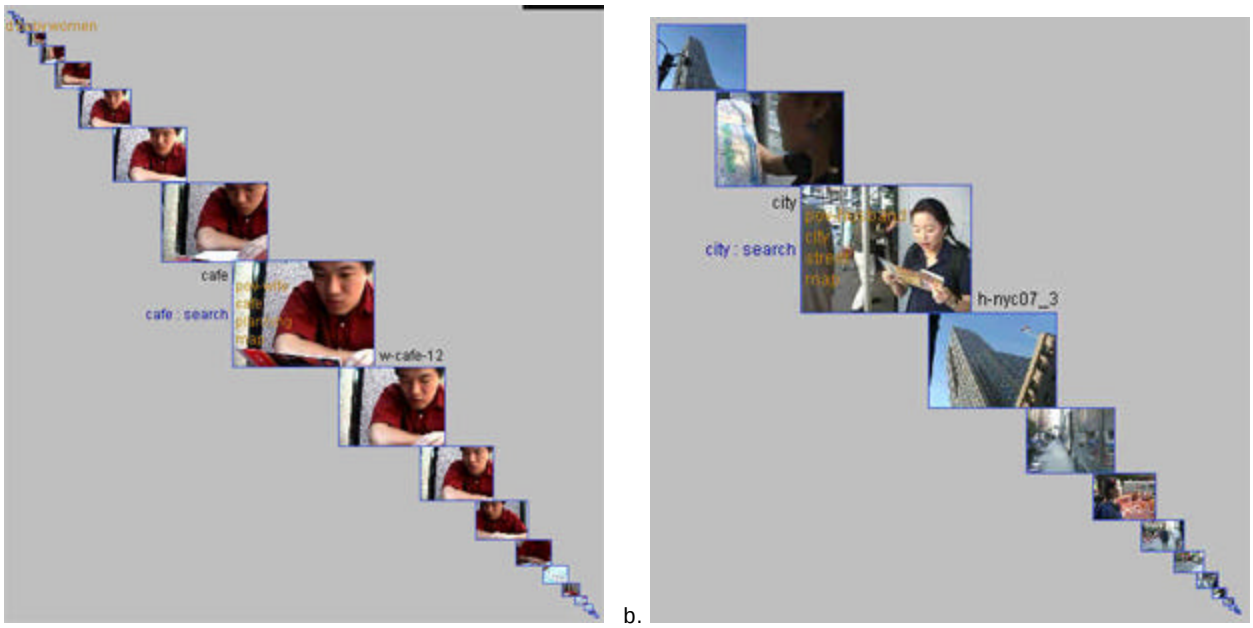


Figure 5-10. a. My "cafe" clips. b. Phillip's "city" clips.

I looked for a way to react to his clips in my next shooting session. Some of the shots showed the wife searching for places to visit, poring over a street map, asking for directions, and looking somewhat lost and bewildered. I decided to explore the husband's reaction to her feeling out of her depth in the big city. I shot footage of the couple walking around, the husband holding the map. He complains bitterly about how they are lost and losing precious time. I intercut these clips with Phillip's clips to create a sequence in which the harshness of his words in my footage are accentuated by her naive confusion in Phillip's footage.

Figure 5-11. The "lost" sequence, titled "they are lost in the city", containing selections from both my "cafe" clips and Phillip's "city" clips.

5.6 Observations

Here are some observations about the Individeo interface, gathered during my experience with the Honeymoon production. While the interfaces may have been well-suited to the tasks they were designed for, there were missing interface elements and features that I found desirable in an actual production. Problems and ideas for future research are described with respect to both the browser and the editor.

Visual quality of the interface

The resolution and texture of Individeo's visual interface suffered due to the technologies used to implement it. Because Individeo was conceived as an online application, I constrained myself to using the Java Development Kit (JDK) 1.1, a programming toolkit which is supported by the most popular Web browsers available today. This way, I could count on Individeo being accessible to the largest possible number of

online users. Consequently, I relied on the JDK's Abstract Window Toolkit (AWT) to manage and render all of my thumbnail images. The AWT in JDK 1.1 does not offer flexible and powerful controls over imaging, unlike later Java technologies such as Java 2D. The result was that the visual interface achieved lower resolution than desired.

Individeo Browser

Once Phillip and I had shared numerous clips and sequences, the Individeo browser was effective in showing me the different contexts in which clips had been reused. After selecting a clip, I could review all sequences containing the clip that we had edited. In addition, I found other uses for the browser that I had not considered in the design phase. As previously described, I used the browser's display of sequences to organize similar clips into groups. This way, while reviewing a sequence, we could also easily see what other alternative clips might work well within that sequence. Another instance where the browser proved effective was when I wanted to edit alternative version of the same sequence. I could compare these versions by scrolling through one of the sequences, and seeing how it was similar or different to another sequence.

The browser was less effective for reviewing brand new footage, when I wanted to look at new clips from Phillip or myself in order to begin making new sequences. This was largely because of the inherent limitation of trying to represent the temporal information inside a video clip using a thumbnail image. A single, static image is hardly a rich representation of time-based content, showing as it does what is visible in only one single instant. In Honeymoon, I often produced many clips that had more or less the same visual content, but different aural / dialogue content. When I imported these clips into Individeo and retrieve them using the keyword search, the browser showed a large number of similar thumbnails, which made it difficult to tell clips apart:

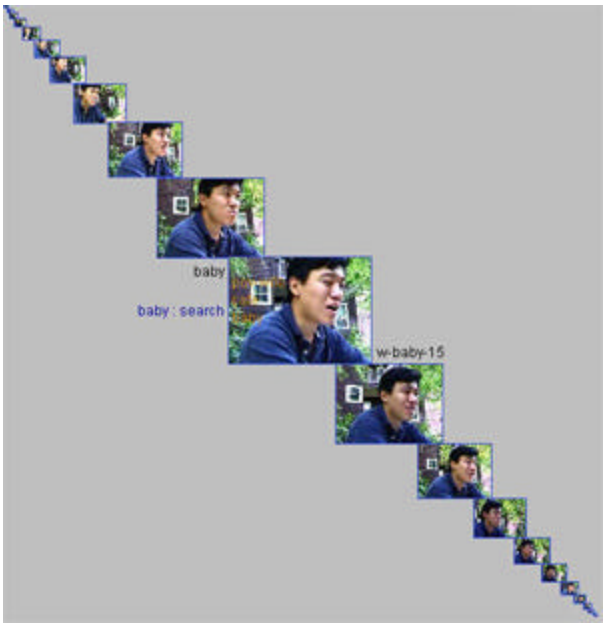


Figure 5-12. Search result for "baby". Because the clips are similar visually, it is difficult to tell the clips apart.

A more traditional interface for listing clips and sequences may be highly complementary to the Individeo browser. Currently, when the application launches, the browser can be configured to display no sequences at all, or to show the sequence with the earliest edit date. To retrieve sequences, the user must first conduct a keyword search for one of the clips in those sequences. Phillip felt that this was frustrating: "It is difficult to know where to begin, as the sequences you have put together have been hard for me to find." To alleviate this problem, a more traditional linear listing of available sequences might complement the browser. A listing could give the user direct access to the most recently edited sequences. Selecting one of the sequences would bring it up in the Individeo browser, along with the other related sequences.

Individeo Editor

With the Individeo editor, I could easily add, remove and reorder clips while editing a sequence. As discussed in previous sections of this thesis, in other timeline-based video editing tools, such actions require some care and precision in placing the clips in the correct position. I found the increased freedom of arranging the clips in the sequence space to be more conducive to trying out various ideas for editing a sequence. Precise measurement of how Individeo's casual editing interface helps users will require more extensive testing.

A significant problem with the editor was that the clip was the atomic unit for editing: once Phillip and I had edited our footage and made our clips, we could not edit those clips further in Individeo. This was a source of frustration for Phillip: "the edit points [in and out points specifying the beginning and end of a clip] I want are pretty different from the edit points you give me. It is interesting in that even though we are "collaboratively" editing, we have to guess what each other wants and there is no way (within the system) to specify what we want." While my view is that being forced to work with each other's edit points is part of the challenge of a collaborative production like Honeymoon, the ability to define custom edit points specific to the sequence seems like a natural next step for the editor. When defining custom edit points is made possible, it would also be important to represent this information in the browser. Each clip in the browser could indicate the existence of edit points and how much of the clip was actually used.

The text inter-titles supported by the editor proved to be less useful for a project such as Honeymoon. Because Phillip and I were focused on editing video-only sequences that relayed some part of the story, it was not appropriate to interrupt sequences wit h text inter-titles. It would have been more useful to be able to add text comments as an additional layer, and to attach these comments to specific sections of a sequence. For example, I might be able to attach a comment such as "these three clips work well together" to a segment within a sequence, beginning at the second clip and ending at the fifth clip. For the production, we relied on other forms of text communication such as emails and a shared weblog to discuss story ideas and production issues. I realized that the text editing and presentation capabilities of Individeo would have to be more extensive if it is to support messaging and discussion between involved parties.

5.7 Summary

Honeymoon is an experimental collaborative video production, where two geographically separated videomakers attempted to tell a story together, by browsing and editing shared video with Individeo's custom software interfaces. This chapter outlined the basic story, framework for collaboration, and

particular examples of how Individeo was used during production. It also presented observations on the strengths and shortcomings of the Individeo interfaces.

In summary, I revisit briefly the design issues for Individeo, described in chapter 4. Here are additional observations from using Individeo for the Honeymoon production, in the context of the design issues.

Shared contexts

The Individeo browser made clear how both Phillip and I were reusing the video clips that each of us had shot and contributed. It accomplished this by visualizing the different sequences and contexts that the clips was edited and placed in. Being able to see these multiple contexts in one unified view made it easier to compare and contrast our creative approaches in terms of editing and story direction.

Scales of construction

The Individeo browser displayed all sequences that share certain clips, and each sequence I edited was shown as linked to other previously edited sequences. My own creative output was always shown as integrated with Phillip's, based on the fact that we were sharing the raw materials. Although the production did not produce as large a database of clips and sequences as I would have preferred, we could still watch a web of interconnected stories growing with every sequence we created.

Casual editing

With the Individeo editor, I could try out different editing options quickly by casually changing the layout of clips. This was especially useful in Honeymoon: since I did not produce all of the content myself, I was seeing many clips for the very first time when I sat down to edit, without preconceived notions of what the sequence might look like . As a result, I was constantly trying out different clips in various configurations for any given sequence.

Communication with media

The Individeo editor supported the mix of text and video within a sequence, where the text would appear as inter-titles. This feature was less applicable in Honeymoon, since we wanted to tell the story with strictly video-based sequences, without seeming interruptions of text clips. The ability to add text comments as a separate layer from video may be more suitable for a fictional production like Honeymoon.

Honeymoon was an opportunity to evaluate Individeo within an actual collaborative video production. The framework for Honeymoon was specially designed to foreground the sharing and reuse of video between the collaborating videomakers. This made it an ideal project to help evaluate the interface design in Individeo, whose aim was precisely to support such collaborative activity. Although Honeymoon was a very specific instance of how Individeo can be used, it allowed me to discover the strengths and shortcomings of the interfaces up close and personally.

6.0 Conclusion

This thesis explored the design of visual interfaces for supporting collaborative video production, and evaluated custom video browsing and editing interfaces through a constrained video production project. The thesis research consisted of two main components:

- Individeo is a networked application for collaborative video editing and sharing. Individeo was designed to provide contextual information about how multiple editors are reusing a set of shared video media. By seeing directly how other users are editing with the shared pool of content, each editor may gain a better understanding of the overall creative process. Individeo also sought to simplify the traditional video editing interface by using a storyboard mechanism, which is especially useful when brainstorming about different options as to how to edit a sequence. Four design issues were raised during the design and development of Individeo: shared contexts, scales of construction, casual editing, communication with media.

- Honeymoon is an experimental collaborative video production, involving two teams of geographically separated directors and actors. Honeymoon was a story about a newlywed couple, told strictly through the points of view of the two characters. The makers of Honeymoon shared and edited their video footage using Individeo. The production helped me in evaluating the effectiveness of Individeo in supporting an actual collaborative production. The four design issues considered during the development of Individeo were reviewed in the context of the production.

The thesis satisfied my goal of using my tools to produce content that I felt strongly about. Building a tool then using it to tell a story is indeed a tradition at Interactive Cinema, and this thesis continues that tradition.

This thesis research yielded the following accomplishments.

- The thesis research produced a *visual interface* for contextual browsing of shared video content. If multiple editors are editing with a shared pool of video, the interface developed through this thesis research can provide a clear picture of how the shared video is being reused by multiple editors, and what different contexts a video clip belongs to as it is edited into various sequences. This information is valuable in a collaboration scenario because each participant can become more aware of the creative activity of others, and because it can be a springboard to other ideas for editing with the shared clip.

- The thesis research conducted a video *production project* with an experimental framework, where two geographically separated videomakers would tell a story by reacting and improvising with and against each other's produced footage. The project was used as the evaluation for the online video interfaces.

- The thesis research identified four *design issues* to consider when building an application to support online collaborative creation: how to visualize the shared contexts among users in terms of the media being shared among them; how to enable construction at different scales, so that each composition becomes both an independent creation as well as a part of a larger whole; how to enhance brainstorming during media composition; and how to support communication and discussion using a combination of text and other media.

It should be noted that Individeo can be used in a variety of other scenarios and applications, two of which were sketched briefly in chapter 2. It may be useful to consider here the question that a constrained evaluation like Honeymoon can only tackle indirectly: what would a large community of users do with Individeo, as they edited and shared video online? How could Individeo be used by a large user base, and what would it mean in terms of media, culture, and social communication? Here I discuss three possible applications for Individeo and related ideas for future research and development of Individeo.

Shared knowledge

When I originally began thinking about possible applications for shareable media, I began with the notion of a video-based newsgroup. Today, news websites such as CNN (http://www.cnn.com/) regularly provide streaming video clips for the day's top stories; these are short snippets from the cable television news reports. In other areas of the CNN website, visitors can discuss the news with others. These areas are in the form of text-based discussion boards resembling a typical newsgroup, where visitors can post and reply to text messages. The video clips and the discussion boards are separate from each other. If a user wants to refer to a video clip in her posting to the discussion boards, she does so by saying "that point in the video where..." and so on.

With Individeo, a news website could provide video content in short chunks, which could then be selected and edited by the visitors as they discussed the news. After watching a sequence edited by the news organization, a viewer could select the particular clips she wishes to talk about, and compose a new sequence with those clips. She could use the text inter-titles to highlight sections of importance in a video clip, support or dispute the facts presented by the clip, and express her opinion about the content of the clip. She might even upload her own video clips. She could shoot herself, webcam-style, as she states her point out loud. Or she might have access to footage that fills a gap left by the original news sequence, providing additional information about the story in question. As the viewers actively contributed their own sequences, discussing the news and sharing their opinions, one would see the growth of a shared pool of various viewpoints and collective knowledge about the news at hand. The community of active viewers would take on the role of information provider, supplying new information and interpretations.

John Seely Brown, in his book The Social Life of Information, discusses shared knowledge as being more than a collection of bits of information. He observed the way a company's workers used stories as a means of sharing group knowledge. When the workers spent time with each other, they would swap tales of their experiences in the field, stories about dealing with customers, products, equipment. These stories would pass among different workers, and would become part of the worker's shared knowledge. Brown writes [Brown 2000]:

> Shared knowledge differs significantly from a collective pool of discrete parts. In this pool of knowledge, where one person's knowledge ends and another's begins is not always clear... It was a collective process that created an indivisible product. Thus we tend to think of knowledge less like an assembly of discrete parts and more like a watercolor painting. As each new color is added, it blends with the others to produce the final effect, in which the contributing parts become indivisible.

Shared knowledge, for Brown, is a collection of information where each piece is linked to others, and there is a common framework for understanding how the pieces are related to each other. A strength of the Individeo browser is in visualizing this type of shared knowledge. In the news scenario, each sequence edited by a viewer would be shown in the browser as linked to other sequences by the connection of common clips. Each new contribution would be seen as a part of the growing body of content, rather than as an independent, discrete addition. The interface enables viewers to explore the collective content through a visual representation of shared knowledge.

Grassroots creativity

The Honeymoon project involved two videomakers engaged in the act of reusing and recontextualizing each other's footage. One can imagine a larger online community using Individeo as a place to share and play with video. Given a video clip, any Individeo user can combine it with other clips to subtly affect or outrageously change the perceived meaning of the clip. In one extended example in chapter 2, a romantic moment from the classic film Gone With the Wind undergoes a change of tone when it is juxtaposed with exploding fireworks, a combination which brings a humorous flavor to the sequence. Another technique for

recontextualizing video would be to separate the images and the audio, and substituting other visuals or audio. While this is currently not possible with the Individeo editor, enabling the separation and recombination of video and audio tracks would not be difficult to implement. This would add flexibility in how a user can recombine heterogeneous footage in various ways.

The art of mixing and matching existing content to create something new is not a new phenomenon, but digital media and software has made it more accessible. If we focus on the medium of moving images, found-footage filmmaking and fan video art have been around for some time, and experienced practitioners have produced highly sophisticated works. Networked tools such as Individeo can provide a wider audience for such works, and also invite the audience to participate in videomaking. Individeo can support the type of non-professional creative endeavor that Henry Jenkins describes in Textual Poachers. In his discussion on fan video art, Jenkins describes the "fan aesthetic" as one that "centers on the selection, inflection, juxtaposition, and recirculation of ready-made images and discourses".' When fan videos mix imagery from a TV show with unrelated pop songs, the lyrics "amplify, critique, or parody" the visuals and storyline. With Individeo, the community of media fans could contribute, select, edit and share their beloved content. They could recombine video, audio and text in order to "amplify, critique, or parody" the original content. As Jenkins writes of fan video art, "the same images, the same shots resurface in video after video, achieving different yet related meanings in each new context; the fascination is in seeing how different artists assign their own meanings to the raw materials they share." The description could apply to the pool of shared content that might grow out of the participation of grassroots media fans and amateur videographers.

The dream of freely sharing and editing mass media content from movies, television shows and commercials raises the issue of copyright for the content. Given the current legal and financial climates, it seems unlikely that copyrighted content would be allowed to exist within the Individeo environment, to be viewed and edited freely by all comers. In one of my conversations with Henry Jenkins, he stated elegantly the parallel between Napster and Individeo. On one hand, with Napster, the user can copy the media file onto their local machines, and modify the actual bits of the content; in Individeo, the media is only streamed and therefore not downloaded. On the other hand, Napster only supports the search and transfer of media files, while Individeo can be used to compose new content (video sequences) which can affect how the original media is perceived. For example, with Individeo, a user cannot download and save a copy of a clip from a

Disney animation, but she could make a less-than-wholesome sequence in which the Disney clip is juxtaposed with violent imagery. The user does not have access to the actual bits of the media, but is able to modify the integrity of the content in terms of its perceived meaning. Such recontextualization will likely incur the wrath of the copyright watchers.

Copyright in the realm of digital media is a complex issue, and this is far from an effort to paint a full picture of the current state of digital copyright. It should be noted that Individeo and Interactive Cinema's Shareable Media project suggest a model for how copyright might work for sharing original media. The authors of the Shareable Media system have proposed an "open source" model for content, based loosely on the open source programming movement. Once a media clip has been submitted to the Shareable Media system (and can be viewed and edited with using Individeo), the media clip can be freely used by all members of the Shareable Media system. No user would have exclusive rights to any of the media clips or sequences. With this model, the community of users could freely share and compose with each other's media.

The difficult question involves the case where a user submits media that is copyrighted, such as a Disney animation clip. While I am not qualified to draft a proposal for how this might be handled, I have two personal thoughts on the issue of circulating copyrighted material within Shareable Media. First, allowing media consumers to play with the media might be the path to greater engagement. Companies that own the copyright to the majority of mass media content have often guarded their content zealously, shutting down fan websites and file sharing services. However, it is my belief that when you can play with content - by modifying the bits directly or creating new context and perceived meaning - you create something of your own, and in doing so, draw closer to the content. If media companies were to actively support project such as Individeo and Shareable Media, it might help the media companies in terms of attracting and holding onto fans of mass media. Secondly, regardless of what the companies do, I believe it is vitally important for a space such as Shareable Media to exist. An increasing amount of mass media is being produced and controlled by a handful of powerful corporations with global reach. There is growing consolidation of control over production and distribution of content. An outlet for independent voices, where small-time producers and grassroots consumers could express their views and exercise their creativity,

would make for a healthy counterbalance to this trend of global consolidation. The Shareable Media applications have the potential to become such an outlet.

Mixed-media messaging

Within a large community of users, Individeo could be used as a tool for communicating with a mix of video and text. The use of video for everyday communication has seen mixed success: while video phones have never taken off in the consumer market, video conferencing is common in the business world. On the Web, there is already a community of webcam users, who publish snapshots or stream videos of themselves, mostly typing away at the computer. One popular form of online communication is instant messaging: users are able to send text messages to each other in real-time and hold conversations. Individeo could be an attractive communications tool for the fans of webcams and instant messaging. Using the storyboard editor, users can quickly create and fire off short messages containing video clips and text. Because the Individeo browser updates automatically when a new sequence has been uploaded, a real-time conversation could be carried out with Individeo. More asynchronous modes of communication can also be supported, since the sequences are stored and archived. Video newsgroups or video weblogs could be implemented using Individeo.

The Individeo browser would be useful in identifying popular topics of discussion among users, as indicated by what content is being reused most actively. In newsgroups, the subject headings of the message or the thread can help the reader understand what is being discussed. The number of messages and their hierarchical organization reveal how popular a topic is. In the case of instant messaging and weblogs, it is difficult to understand the hot topic without scanning the majority of the text. Trying to understand the nature of distributed communication is an active research topic, addressed by projects such as Blogdex (http://blogdex.media.mit.edu/) at the MIT Media Laboratory. If Individeo were used for mixed-media messaging, each message would contain video and text, and a video clip that is the subject of heated discussion will likely appear in numerous sequences edited by many different users. When the popular clips is chosen as the focus clip in the Individeo browser, a large number of associated sequences will be displayed. By browsing and looking for popular clips, a user can quickly identify which clips are the current "hot" content, and which topics are being discussed by a large number of users.

Additional features to the Individeo interface would help support the quick and efficient creation of video messages. Currently, only video clips in a sequence can be transferred from the browser to the editor when creating a new sequence. Text clips created by other users cannot be added to new sequences. If users are carrying out a conversation with text and video, it follows that any part of an existing message should be quotable in a replying message, just as in email messages or newsgroup postings. The simple actions of cut, copy and paste in text editors provide ease and flexibility in composing a text document, and it would be worthwhile to investigate how a video editing interface can supply the same flexibility. Also, users may wish to keep certain clips permanently in the Individeo editor, which they can use repeatedly in different sequences, without having to find them in the browser each time. These clips might be the equivalent of the emoticons we find today in text-based communication, a symbolic element that indicates the mood of the writer and the intended tone of the message. The fireworks clips from the extended example in chapter 2 might be one such clip, used for humor and parody. It seems likely that visually striking yet ambiguous video clips would be widely used in communicating with Individeo, and one further area of study would be to survey what types of footage is fun and easy to make messages with.

References

Aigrain, P.; Zhang, H.; Petkovic, D. Content-based representation and retrieval of visual media: A state-of-the-art review. Multimedia tools and applications, volume 3, pp. 179 - 202. Kluwer Academic Publishers, 1996.

Berners-Lee, Tim. Weaving the Web. New York: HarperCollins, 1999.

Bordwell, David. Narration in the Fiction Film. Madison: The University of Wisconsin Press, 1985.

Bresson, Robert. Notes on the Cinematographer. England: Quartet Books, 1986.

Brown, John Seely; Duguid, Paul. The Social Life of Information. Boston: Harvard Business School Press, 2000.

Elliott, Edward. Watch * Grab * Arrange * See: Thinking with Motion Images via Streams and Collages. MIT MS Thesis, 1993.

Evans, Ryan. LogBoy Meets FilterGirl: A Toolkit for Multivariant Movies. MIT MS Thesis, 1994.

Furnas, George. Generalized Fisheye Views. Human Factors in Computing Systems CHI '86 Conference Proceedings, pp. 16 - 23, 1986.

Halliday, Mark. Digital Cinema: An Environment for Multi-Threaded Stories. MIT MS Thesis, 1993.

Jenkins, Henry. Textual Poachers. New York: Routledge, 1992.

Kelliher, Aisling. PlusShorts: Using Punctuation as an Iconic System for Describing and Augmenting Video Structure. MIT MS Thesis, 2001.

Kelliher, Aisling; Seo, James Jung-Hoon; Pan, Pengkai; Lin, Cathy; Davenport, Glorianna. Visual Interfaces for Shareable Media. International Symposium on Electronic Art, Paris, 2000.

Kilian, Axel. Defining Digital Space Through a Visual Language. MIT MS Thesis, 2000.

Lamping, J.; Rao, R. The Hyperbolic Browser: A focus + context technique for visualizing large hierarchies. Journal of Visual Languages and Computing, volume 7, pp. 33 - 55, 1996.

Lin, Cathy; Cheng, David; Pan, Pengkai; Seo, James Jung-Hoon; Kelliher, Aisling; Davenport, Glorianna. An Extensible Architecture for Multiple Applications with Shareable Media. International Workshop on Networked Appliances, New Brunswick, 2000

Mackinlay, Jock; Zellweger, Polle. Browsing vs. Search: Can We Find a Synergy? CHI '95, Denver, 1995.

Maeda, John. MAEDA @ MEDIA. New York: Rizzoli, 2000.

Pan, Pengkai. I-Views, a Storymaking Community of, by and for the Audience. MIT MS Thesis, 1999.

Ponceleon, Dulce; Srinivasan, Savitha; Amir, Arnon; Petkovic, Dragutin; Diklic, Dan. Key to Effective Video Retrieval: Effective Cataloging and Browsing. ACM Multimedia, Bristol, 1998

Smith, Thomas Aguirre. If You Could See What I Mean... Descriptions of Video in an Anthropologist's Video Notebook. MIT MS Thesis, 1992.

Yeo, Boon-Lock; Yeung, Minerva; Wolf, W.; Liu, Bede. Method and Apparatus for Video Browsing Based on Content and Structure. United States Patent, number 5708767, 1998.

Yeung, Minerva; Yeo, Boon-Lock; Wolf, W.; Liu, Bede. Video browsing using clustering and scene transitions on compressed sequences. Proceedings of Multimedia Computing and Networking, volume SPIE 2417, 1995.

Yong, R.; Huang, T. S.; Mehrotra, S. Constructing Table-of-Content for Videos. Multimedia Systems, volume 7(5), pp. 359 - 368, 1999.

Zhong, Di; Zhang, HongJiang; Chang, Shih-Fu. Clustering methods for video browsing and annotation. Technical report, Columbia University, 1997.