Some Assembly Required: Cinematic Knowledge-Based Reconstruction of Structured Video Sequences

by
David José Tamés

B.S., Computer and Cognitive Sciences
B.A., Philosophy
University of Florida
1983

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

June, 1996

Signature of Author _____
Program in Media Arts and Sciences
February 9, 1996

Certified by _____
Glorianna Davenport
Associate Professor of Media Technology
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by _____
Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Some Assembly Required: Cinematic Knowledge-Based Reconstruction of Structured Video Sequences

by
David José Tamés

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, on February 9, 1996 in partial fulfillment of the requirements for the degree of Master of Science in Media Arts and Sciences.

Abstract

Cinema is a stream of audio and visual elements orchestrated in time providing us with a rich language and tradition of storytelling. New media technologies, particularly structured video, open up possibilities to expand and evolve the language of cinema and establish new modes of content delivery.

The work described in this thesis investigates and demonstrates the cinematic potential of structured video techniques. With structured video we are no longer limited to the immutable frames that constrain traditional cinema. We are now able to composite images from a collection of media objects in real time at the moment of presentation.

This thesis describes the design and production of *Two Viewpoints,* a short narrative providing the viewer with a choice of subjective viewpoints. *Two Viewpoints* is implemented using structured video techniques. The design and implementation of an authoring toolkit for developing and previsualizing structured video movies is discussed.

Thesis Supervisor: Glorianna Davenport

Title: Associate Professor of Media Technology

This work was supported by the Television of Tomorrow Consortium

Some Assembly Required: Cinematic Knowledge-Based Reconstruction of Structured Video Sequences

by

David José Tamés

The following people served as readers for this thesis:

Reader _____

V. Michael Bove, Jr.
Associate Professor of Media Technology
Program in Media Arts and Sciences

Reader _____

Michael Lehmann
Director and Producer

Acknowledgments

I am especially grateful to my advisor, Glorianna Davenport, for her guidance, patience and support and for inviting me to the lab in the first place. Thanks to my readers, Michael Lehmann and Michael Bove, for their comments and suggestions. Drew Beechum and Edgar Ngwenya, as UROPs, provided coding assistance during many late night hours.

My fellow students and colleagues at the Media Laboratory have been a constant source of illuminating discussion and questions, thanks for your ideas, conversation and encouragement along the way: Mark Halliday, Ryan Evans, Michael "Wave" Johnson, Stephan Fitch, Tinsley Galyean, Steve Drucker, Warren Sack, Jennifer Gonzales, Sara Elo, Josh Bers, Sherrie Lassiter, Flavia Sparachino, Bill Burling, Eddie Elliot, Scott Higgins, Natalia Tsarkova, Amy Bruckman, Nicolas Saint-Arnaud, Baback Moghaddam, Claudio Pinharez, Kevin Brooks, Mike Murtaugh, Freedom Baird, Michael Massey, the list goes on and on. David Kung, in particular, expanded my musical, televisual, and pop-culture horizons. Betsy Brown, Linda Peterson and Santina Tonelli were always there to help navigate the bureaucratic labyrinth.

I have been fortunate over the years to have been influenced, inspired, and guided by wonderful teachers, friends and angels including my parents and close family, Patricia Thar, Celia Lighthill, Harry Mathias, Debbie Brubaker, Caroline Blair, Dave Scardina, Zev Berman, Cheryl Buchinski, Ted Olsson, Tom Haddock, Mark Goddard, Kathryn Blair, David Beauchamp, Larry Feuer, Eric Oesterle, Donna Cunningham, Harold Hedelman, Henry Jenkins and Martin Roberts. Much of what I've accomplished as a cinematographer and media technologist I owe to their teaching, encouragement, and support. Special thanks to Flavia Sparacino for words of encouragement at just the right moments, Natasha Lofgren for her friendship, Peter Rea for making me laugh at my moment of greatest despair, and Christine Parker for her positive energy.

I owe a special gratitude to the directors with whom I've collaborated: Julie Chang & Joe Rubin (*Slack Trek*), Tony Romain (*The Yellow Wallpaper*) and Stephen Kijak (*Never Met Picasso*). These projects have helped me grow artistically as a cinematographer. I learned so much working with the *Two Viewpoints* team: Shawn Becker, Stefan Agamanolis, Araz Inguilizian, David Kung, Katy Brown, Roger Yergeau, John DeSimone, Roger Marbury, Chris Benda, and Michael Rosenbush, John Watlington, Bill Butera, Eugene Lin, Andrew Beechum, Edgar Ngwenya, Tony Romain, Rachel Lillis and Michael Slayton.

The right music made many things possible during the course of my thesis work: Dead Can Dance, Pylon, Talking Heads, Unmen, System 7, Aphex Twin, The Primitives, The Breeders, B-52s, The Orb, Kristin Hersh, Liz Phair, Joe Jackson, John Coltrane, Bettie Serveert, Philip Glass, Brian Eno, Tom Tom Club, Stereolab, and Tangerine Dream.

Cinematic interludes put things in perspective while I was writing: Atom Egoyan's *Exotica,* André Techine's *Wild Reeds,* Denys Arcand's *Love and Human Remains,* Lars von Trier's *The Kingdom,* Oliver Stone's *Natural Born Killers,* Derek Jarman's *The Last of England,* Wim Wender's *Notebook on Cities and Clothes,* and Daisy von Scherer Mayer's *Party Girl.*

Contents

1 Introduction 9

Structured Video 9

- Cheops 10
- With structured video the frame becomes dynamic 10
- Structured video enables interactive and multivariant stories 10
- Passive or Active Viewers? 10
- Structured video enables an infrastructure for semantic annotation 10

Two Viewpoints 11

The SAR Toolkit 11

- A tool for designing and thinking about multivariant stories 12
- SAR provides a framework for encoding the maker's intent 12

2 Background 14

Narrative 14

- Narrative is a fundamental perceptual activity 14
- Narrative is an object and narration a process 14
- (de)construction? 14

Cinematic Narrative 15

- Cinema History, Aesthetics and Narrative 15

Interactive Cinema 16

- Interactive cinema expands our notion of cinema 16
- Constraint-based cinematic editing introduces the computer as a computational partner in the storytelling process 16
- Fluid interaction helps preserve narrative reverie 16
- Integrating story breaks with the narrative improves viewer engagement 17
- Descriptive story structures facilitate development of multivariant movies 17
- Good representation facilitates reuse of media objects 18
- Narrative Intelligence? 18
- Computational partners must orchestrate sound & picture asynchronously 18

- Intelligent cameras and narrative guidance help provide a fluid experience 18

Structured Video 18

- Whither HDTV? 18
- Structured video represents video sequences in terms of component parts 19
- Structured video enables the development of new story forms 19

Point of View and Narrative 20

- Constraints, guidance & subjunctivization 21
- The problem of person in narrative 22
- Lady in the Lake 22
- Voyeur 22
- Creating subjectivity 23

Little Dorrit 23

- Little Dorrit's Story 23
- He Said, She Said 24

The Cinematographer's Activity 24

- Perceptually, we may use either an inductive or deductive approach when making sense of an image depending on the size of the image 25
- Cinematographers today often have to compose for multiple aspect ratios 27
- Blade Runner and T2 have been transferred to video with different results 29
- Rent some videos, make some popcorn and take a look for yourself 29

Master Scene Cinema Language 30

- Master Scene Language in practice 30

3 Two Viewpoints 32

Objectives 32

- Explore production techniques for structured video and Cheops 32
- Not the Royal We 32
- The Viewers Experience 33

Not just sequences anymore... 34

The evolution of Two Viewpoints 34

- Adapting the story 34
- Gilman's The Yellow Wallpaper 34
- Romain's Adaptation 34

Production 35

- The Set 35
- Capturing the Actors 35
- John as seen from each camera 36
- Two Viewpoints Set Plan 36
- Calibration Cubes 37

- Camera Views 37
- Calibration 37
- Production Team 38

Postproduction 38

- Video objects 38
- Digitizing Video, chromakey and segmentation 39
- Orchestration 39

The Viewer's Experience 39

- Playout from John's Perspective 39
- Playout from Kathy's Perspective 40

4 The SAR Toolkit 41

Motivation 41

Scope 41

Design Approach 42

The Moviemaker's Activity 42

The Toolkit 43

- The SAR toolkit 43
- Shot Composer 44
- Sequence Orchestrator 44
- XmDat 45

Implementation Notes 46

- Structured Video Objects 46
- Shot Composer 46
- Sequencer Orchestrator 46
- XmDat 46
- An Integrated Toolkit 47
- Performance 47

5 Evaluation, Reflection, & Suggestions for Future Work 48

Two Viewpoints 48

- Interaction must be integrated with the narration 48
- Hardware limitations 48
- The frame rate, albeit low, is not a major problem 48
- Is it really a multivariant movie? 48
- How 'smart' is it? 49
- Alternative Interface 49

- Analog video blues 49
- Characters did not blend into the scene in a seamless manner 49
- Actors and objects did not cast shadows 50

The SAR Toolkit 50

- Export Isis scripts 50
- Extensibility 50
- Performance 50
- Support the use of sound 50
- Provide more intelligent assistance to the user 51
- More flexibility in the playout algorithm 51

Structured Video Production Methods 51

- Pre/Pro/Post 51
- Postproduction 51

Reflective Moments 52

A1 TwoViewpoints Original Scenerios 53

John 53

Kathy 53

Interaction 53

A2 Two Viewpoints Original Script 54

Kathy's Perspective 54

John's Perspective 55

A3 Two Viewpoints Shot List 57

A4 Blue Screen Compositing 58

References 60

1 Introduction

Cinema is a stream of audio and visual elements orchestrated in time providing us with a rich language and tradition of storytelling. Recent research at the Media Laboratory emphasizes the personalization and dynamic presentation of content (Davenport, 1995). New media technologies, particularly structured video (Bove, Granger and Watlington, 1994), open up possibilities to expand and evolve the language of cinema and establish new modes of content delivery.

The work described in this thesis investigates and demonstrates the cinematic potential of structured video techniques. With structured video we are no longer limited to the 'atomic' media elements with immutable frames of traditional cinema. We are now able to composite images from a collection of media objects in real time. I describe the design and production of a short narrative providing the viewer with a choice of subjective viewpoints, implemented using structured video techniques. In addition, I explore the design and implementation of an authoring toolkit for developing and previsualizing movies made from structured video components.

Structured Video

Structured video as a technology enables the separation of the various media objects that make up an image during production and postproduction. Separate media objects are assembled together in real time at the moment of presentation to create the final image. Shots produced using structured video are independent of a specific aspect ratio, framing, and screen size. Structured video objects can be annotated with data referring to their physical characteristics, syntactic relation to other objects, or their semantic relationship to other objects and higher level concepts. Using structured video techniques enables several new modes of content production and reception. For example, a multivariant movie may designed with structured video techniques in order to play out differently based on viewer interaction.

Cheops

In order to demonstrate and experiment with structured video techniques, the Media Lab has developed Cheops, a compact, modular platform for the acquisition, processing, and display of digital video sequences and model-based representation of scenes (Bove & Watlington, 1995).

With structured video the frame becomes dynamic

Filmmakers frame, shoot, and direct scenes to tell a specific story. Yet in the end the only thing that makes it through the distribution pipeline is a tiny trickle of image and sound bits. The information which motivated the decision process behind the meaning of the bits is lost. Today most filmmakers have control over the framing of a film only during its initial theatrical release. Films are projected on large wide screens, while videos are played back on televisions of different shapes and sizes. Computer window systems cloud the scene with their ability to resize windows. In the future you may have a high-resolution large screen projection television in your home with high-fidelity surround sound, however, while on vacation you may want to catch the next episode of your favorite show on a small hand-held device. A structured video representation makes these applications possible.

Structured video enables interactive and multivariant stories

Remote controls and the proliferation of channels have introduced a simple notion of television interaction through channel-surfing, however, cinema and television remain fundamentally passive, one-size-fits-all media. A creative application of structured video is delivering a multivariant movie in which both the composition of the shots *and* the sequencing is dynamically controlled and rendered at the display.

Structured video enables an infrastructure for semantic annotation

In this thesis we begin to explore how it could be possible for the creator's artistic motivations and intentions to survive through the process of delivering content. The content may be designed to play out in a multivariant fashion or framed in different ways depending on the specific playback device.

As decisions are made throughout the production process, some cinematic knowledge should be encoded with the component parts of an image. This knowledge includes the cinematographer's use of framing to express character identification and perspective; the editor's sequencing of shots based on emotional and rhythmic concerns; the director's preference for a particular actor's performances, etc. The creation of multivariant media challenges us to develop ways to encode

A Century of Cinema

December 28, 1995 marks the 100th anniversary of commercial cinema. On this day, one hundred years ago, the Lumière brothers showed the cinématograph for the first time to a paying public for an admission charge of one franc (Toulet, 1988).

Cinema, a product of 19th Century technology, was born from the illusion of movement by the projection of a series of still photographic images on a screen. New techniques such as structured video take cinema beyond it's mechanical origins.

Passive or Active Viewers?

John Fiske (1987) argues that cinema and television are not 'passive' mediums and that viewers are sophisticated readers of cultural messages and play an active role in the construction of meaning. I agree, however, within the context of this thesis I'm using *passive* in the sense of the viewer's involvement in the process of narration (discussed in the background section).

some of this information along with the media objects.

Two Viewpoints

In order to demonstrate the narrative potential of structured video in general, and Cheops in particular, I produced *Two Viewpoints* in collaboration with Shawn Becker, Stefan Agamanolis, Araz Inguilizian and Tony Romain. *Two Viewpoints* is a short structured video narrative currently running on the Cheops imaging system. We designed the narrative independent of a specific shot sequence, aspect ratio, framing, and screen size.

The actors were shot against blue screen in order to provide maximum flexibility in the final framing and blocking of the scene. Framing and editing decisions, as well as the images required to convey that story, have been encoded in the system with the video objects. This provides Cheops with the 'cinematic knowledge' necessary to reassemble the story in the appropriate manner along two dimensions: (1) the proper framing and sequencing for a particular display size, and (2) the appropriate shot selection to provide the user access to their selected point of view within the story.

"We live lives based upon selected fiction. Our view of reality is conditioned by our position in space and time—not by our personalities as we like to think. Thus every interpretation of reality is based upon a unique position. Two paces east or west and the whole picture is changed"

— Lawrence Durrell, *Balthazar*

In *Two Viewpoints* the viewer's empathy may be aligned with either the character of Kathy or John. Kathy is a writer who does not pay much attention to the 'real world' around her. John, Kathy's husband, is somewhat annoyed by her lack of attention to domestic life. Depending on the viewer's preference, the scene is sequenced and framed dynamically and in real time by Cheops to represent either Kathy's or John's perspective.

Two Viewpoints provides an opportunity to experiment with the subtle and fluid nature of point of view. Traditional filmmakers are very aware that the choice of framing, blocking and performance effect the shape of an entire scene. The vocabulary and syntax of linear cinema is well established and filmmakers are able to learn from the cinematic tradition as well as from each other. Creating a multiple point of view narrative, on the other hand, offers a tremendous challenge to the moviemaker who needs to experiment with new techniques yet still provide the viewer the specific framing and sequencing that conveys the story through each character's perspective.

The SAR Toolkit

Creating a structured video sequence in which the framing is dynamic and the playout multivariant requires a rigorous

approach to design and production. Within a structured video framework, any shot becomes possible. In order for the computer to be an effective computational partner with the director, editor and cinematographer, some form of story knowledge, and the images required to convey that story, must be encoded to provide a playout engine with the necessary parts and instructions to reassemble the story in a manner that preserves the moviemaker's intentions. With the **SAR Toolkit** we demonstrate that even simple, sketchy annotations describing 'coverage' for a scene can enhance the ability of a playout engine to deliver an effective experience.

A tool for designing and thinking about multivariant stories

The **SAR Toolkit** provides a preliminary step towards a tool that facilitates designing and thinking about multivariant movies. As a cinematographer, I am particularly interested how point of view is created in a cinematic context as well how the scene will appear on a variety of display devices. The **SAR Toolkit** allows the director, cinematographer and editor to define synthetic shots and camera movements, as well as narrative constraints. Annotations can be made to customize the composition relative to display size. This system points towards better integration between preproduction, production, and postproduction. The **SAR Toolkit** allows the moviemaker to construct shots and sequences using a visual interface rather than specifying the shots and sequences using a scripting language such as **Isis** (Agamanolis, 1996).

SAR provides a framework for encoding the maker's intent

Ideally the filmmaker should control the aesthetic decisions of framing the movie for different display environments. This could be done with annotations and a set of rules for a renderer to follow. As a cinematographer, I carefully decide with the director what objects appear in or out of the frame and the relationship the objects have with the visual 'code' of the film. Preservation of artistic intent is contingent upon the knowledge and visual elements used to deliver a specific story.

I'm horrified by the butchering that is done to many widescreen films as they make their way into the home via video. Most theatrical films are framed with a 1.85:1 or 2.35:1 aspect ratio, while television provides a 1.33:1 ratio. Although it may be preferable to see a movie in the original aspect ratio, we lose more than aspect ratio when a scene is cropped, especially on smaller display devices. The audience is deprived of the creator's artistic intentions when essential elements of the story are cropped or shown so small that they become unrecognizable. In order to change this situation, some method of anno-

Preserving narrative coherence

In *Blood Simple* (Joel & Ethan Coen, 1985), Visser, a sleazy private detective, is hired by Marty, a saloon keeper, to murder his adulterous wife and her lover. Visser sees this scenario as an opportunity to murder Marty and frame the wife and her lover for the murder. Visser murders Marty but leaves his lighter in Marty's office, the only evidence linking him to the scene of the crime. At one point in the story, several fish are placed on top of the lighter, thus, in later scenes, the various characters who visit the scene of the crime fail to see Visser's distinctive and unique lighter under the fish, which would link him to the scene of the crime. The director has privileged us as viewers with the information that the lighter is there, and this is a source of suspense for several scenes.

If this scene were recomposed for a different aspect ratio or screen size in such a manner that we can't see that Visser's lighter is hidden under the fish, the meaning of the scene would be changed, there would be no suspense. If the storyteller could make some annotations that encode some story knowledge along with the images, then an 'intelligent' structured video decoder would know that this is a critical element and would make sure that the viewer is able to recognize it. On a very small display device this may entail introducing a new cut to a close up shot where there once was none.

While we may have to sacrifice some aspects of the mise en scène, at least the artist would be able to encode some 'hints' as to how the display device could preserve some of the essential elements of the story.

tation needs to evolve. The motivations, techniques, and knowledge behind the scenes in a film are stored in the frames themselves, locked in an indecipherable code. If we can unlock, or store this knowledge, in the stream of bits entering the home, then maybe something could be done to end the slaughter.

2 Background

Narrative

This thesis concentrates on issues regarding the design of a narrative, which we understand as "the recounting of two or more events (or a situation and an event) that are logically connected, occur over time, and are linked by a consistent subject into a whole" (Stam, Burgoyne & Flitterman-Lewis, 1992, p. 69). Narrative is a way of understanding and communicating the meaning of events and the transformative effects of an action, particularly within the context of a *temporal series* (Prince, 1987, p. 60). As we expand our notion of narrative to accommodate new modes of interaction and playout, our most important design goal remains the creation of an experience that is logically connected, occurs over time, and is linked by a consistent subject into a coherent whole.

Narrative is a fundamental perceptual activity

There is something fundamentally human about narrative, it has existed in every known society. Branigan (1992) and Bordwell (1985) describe narrative as a perceptual activity that helps us organize our experience and integrates many aspects of our temporal, spatial, and causal perception. Jerome Bruner describes narrative as one of the two modes of thought. He writes that paradigmatic thinking "attempts to fulfill the ideal of a formal, mathematical, system of description and explanation," while narrative, on the other hand, "deals with human or human-like intention and action and the vicissitudes and consequences that mark their course" (Bruner, 1986, p. 12-13). Schank and Abelson (1995) go so far as to postulates that virtually all human knowledge is based on stories constructed around past experiences and that new experiences are interpreted in terms of old stories.

Narrative is an object and narration a process

Most narrative theorists make a distinction between *narrative* and *narration.* Simply put, narrative is what happens in the story, it is the object or the end result of narration, which is a process. The process of narration regulates and distributes the

(de)construction?

It may be difficult to see the application of the deconstructionist approaches taken in film theory to the construction of content, however, one reason why we take a deconstructionist approach to film theory is because we are forced to. If story knowledge came with the content, we would not need deconstruction techniques. We believe that the motivations and intentions of the director, writer, editor, and other contributors, which now must be conjectured through deconstruction, can be applied towards (re)construction.

information which determines when and how a viewer acquires knowledge from a cinematic experience. Among these processes are the creation of a scene of action, the definition of a temporal progression, and the dramatization of an observer of events.

Something I consider has been disregarded or at least misunderstood by the authors of many 'interactive narratives' has been the distinction between the notions of narrative and narration and their role in narrative sense-making. Branigan (1992) presents a theory of narrative that discusses at length the ideas and concepts related to the process of narration. Branigan states that narration is the overall regulation and distribution of knowledge which determines when and how a viewer acquires knowledge from a text. Branigan's theory is primarily descriptive and desconstructionist, however, since his work is based on a cogntivist framework, I believe that it provides a starting point for a constructivist theory of interactive narration.

I suggest that the important question to ask as we design an interactive narrative becomes what do we reveal, what do we hide, and when, not 'where can we go now' as most branched structured narratives have done to date. In other words, what we are looking for is guidelines for a motivated aesthetic for multithreaded stories that can play out with some user control.

Cinematic Narrative

Cinematic narrative orchestrates a stream of audio and visual elements in time. It is the temporal nature of cinematic narrative that distinguishes it from other narrative forms. It is important to understand how composition, framing, and sequencing of shots influences our perception of a story. These factors play an important role in establishing subjectivity and point of view within a narrative and is discussed extensively in Branigan (1984).

A salient difference between novels, theatre and film is the particular way they articulate point of view. We watch a play as we will, but we can't backtrack. We can read a novel at our leisure, going back or forward and reading or rereading sections as we will. Television provides us with a relentless barrage of options, 24 hours a day. Hypermedia, such as CD-ROMs and the World Wide Web allow us to browse and jump around as we like. We can watch a film in a theater only as the filmmaker wants us to see it; with home video we can fast forward and pause. Interactive and multivariant movies offer us the opportunity to bring together aspects of theatre, novels, television,

Cinematic Syntax

James Monaco (1981) writes, "film syntax must include both development in time and development in space. In film criticism, generally, the modification of space is referred to as mise en scène. The French phrase literally means 'putting in the scene.' The modification of time is called montage (from the French for 'putting together') ... the tension between these twin concepts of mise en scène and montage has been the engine of film aesthetics ever since Lumière and Méliès first explored the practical possibilities of each at the turn of the century."

Cinema History, Aesthetics and Narrative

Monaco (1981) provides an excellent introduction to the technological, economic, political, and aesthetic factors that have shaped film history.

Zettl (1990) provides an illuminating examination of the aesthetic principles of contemporary cinematic and televisual techniques.

Elsaesser (1990) provides a detailed historical perspective on the evolution of cinematic language.

Bordwell (1985) is the definitive work on narrative and narration, starting with a review of previous diegetic and mimetic theories of narrative and then presenting his theory in contrast to these. Narration is discussed with extensive examples of the dominant forms of narration: classic, art-film, and parametric.

Branigan (1992) provides a concise presentations of narrative theory as it relates to film, with good coverage of subjectivity and point-of-view.

Both Bordwell (1985) and Branigan (1992) follow in the tradition of formalist narrative theory and are based on a solid cognitivist framework. Designers of new media forms will find much of relevance in these book.

hypermedia, and film.

Interactive Cinema

For over a hundred years cinema has been a continuous and linear medium. This has been influenced by social, economic, and technological factors (Monaco, 1981). There is no a priori dictum limiting cinema to the physicality of strips of film spliced together in a linear sequence and projected continuously at 24 frames per second. The technologies of interactive digital media, intelligent, high speed computation, and high-bandwidth network communications provide enabling technologies for a new medium of expression that is inherently flexible and non linear in regards to its audio and image generation, processing and manipulation capabilities within an aesthetic context.

Interactive Cinema reflects the longing of cinema to become something new, something more complex and more personal, something which renders an experience dynamically, as if in conversation with an audience. Like cinema before it, interactive cinema reflects not only desire and imagination, but also speculation, technology and commerce.

— Glorianna Davenport

Interactive cinema expands our notion of cinema

The essence of cinematic language—the composition of the frame and the assembly of sequences of shots over time—remain as elemental building blocks, however, in digital environments they are no longer immutable elements. Interactive Cinema adds a new structuring element: variability of the construction of these elements over time, with the possibility of this variability being controlled through a computational partnership, either between the moviemaker and the computer, or the viewer and the computer, or all three.

This flexibility imposes immense aesthetic demands on the moviemaker to create an experience for the viewer that is logically connected, occurs over time, and is linked by a consistent subject into a coherent whole. This issue is fundamental to many past and current research projects in the Interactive Cinema Group, some of which are discussed briefly in the following sections.

Constraint-based cinematic editing introduces the computer as a computational partner in the storytelling process

Rubin (1989) presented a computer-based editing system whose design drew on the principles of cinematic narration. His system automatically generated sequences that conformed to a set of story constraints such as character emphasis and pacing.

Fluid interaction helps preserve narrative reverie

The Digital Micromovie Orchestrator (DMO) provided a framework for sequencing shots in which the computer chose the appropriate clips from a database of annotated clips as the story progressed over time, a process described as orchestra-

tion. *The Endless Conversation* was implemented using the DMO. The orchestration of clips was continuous, allowing the viewer to interact by making selections or just letting the movie play without interaction. This form of interaction is referred to as fluid interaction (Davenport, Evans and Halliday, 1993).

Integrating story breaks with the narrative improves viewer engagement

In contrast to the fluid interaction of the DMO, *Lurker* (Morgenroth, 1995) integrates the interruptions between video sequences into the user experience. Morgenroth tells a story about computer hackers in which he incorporates e-mail messages, Web pages, and video sequences. *Lurker* uses the Internet as a medium for distribution and the basis of interaction. Multiple participants are challenged to piece together the story over a period of days.

Descriptive story structures facilitate development of multivariant movies

Databases of annotated media objects driven by a story template enable us to build multivariant movies which are more dynamic than traversing hard-wired links. Evans (1994) describes two interrelated tools: *Logboy* and *FilterGirl. LogBoy* is used to annotate video clips with sketchy descriptions. Then, by applying story structures, or filters, against these descriptions, *FilterGirl* sequences the clips for playout. These story structures can be dynamic and change over time.

The history of motion pictures is closely tied to the history of the pictorial arts. James Monaco (1981) writes "both cubism and montage eschew the unique point of view and explore the possibilities of multiple perspective." The history of cinema is full of attempts to create multiple perspectives, and even interactive experiences.

Descriptive story structures provide a framework for encoding knowledge about the story and the manner in which it should change for multivariant playout. Evans framed the problem of building a multivariant movie as a process of annotating content and sequencing that content based on attached descriptions. Story constraints are built which take into account narrative structure, cinematic continuity and user preferences. These constraints guide the movie through playout by selecting pieces of digital video based on their descriptions and sequencing them appropriately.

Evans also demonstrated the notion of fluid interaction with *Just One Catch,* a short interactive narrative built with *LogBoy* and *FilterGirl.* The viewer could interact by making choices before the movie played and then *FilterGirl* would build a sequence on the fly based on your choices. The system did not require you to stop and make choices throughout the story, it just kept moving forward in the narrative, however, if you wanted to, you could stop the system, change some parameters, and then watch the movie again.

Good representation facilitates reuse of media objects

Media Streams (Davis, 1995) provides an environment for annotating the physical characteristics of video shots and sequences using an iconic language. A reusable and sharable representation is created that facilitates the sharing of media objects between different people or programs. The system retrieves shots and sequences based on descriptions expressed in its iconic language.

Computational partners must orchestrate sound & picture asynchronously

Kung (1995) reminds us that most video orchestration systems to date have manipulated audio and picture as a single unit. A digital video system should combine descriptive story structures with the ability to manipulate audio and picture separately. Kung describes MILESTONE, a tool for video annotation that creates a knowledge representation framework which is used, in turn, to orchestrate asynchronous playout of sound and picture. Kung presents a series of examples in which the picture varies while the soundtrack is held constant. These examples demonstrate the influence of narration on our interpretation of a cinematic sequence.

Intelligent cameras and narrative guidance help provide a fluid experience

Galyean (1995) suggests an alternative analogy for 'narrative guidance' which he calls 'the river analogy,' suggesting that the audience is in a boat floating down a river, with some latitude and control, while also being pushed and pulled by the predefined flow of the water. For complete flexibility in the control of narration our storytelling system must have the ability to do more than just choose clips from a database of pre-shot video. The system must be able to synthesize new shots to some extent based on the actions of the user within the interactive immersive environment. With *Dogmatic,* a short narrative generated in a virtual environment, Galyean demonstrated that changes in point of view and control could occur without disorienting the user. The system assured that particular plot points were experienced while allowing the user to interact in a continuous manner within story (Galyean, 1995, 101-106).

Narrative Intelligence?

Multivariant movies must contain information on how to construct each shot and how to order the shots so that we can build a sequence that maintains narrative coherence. There must be an overall story structure of some kind.

In chapter 4 we discuss *Sequence Orchestrator,* a tool for annotating alternative shots and cut points in order to enable multivariant playout. We are careful to avoid any claim that this sequencer is doing anything "intelligent." We would prefer to stay clear of the AI problem. This sequencer is not a system that understands stories the ways humans do, however, with it we are able to encode some of our "artistic intent" and provide "templates" for the construction of sequences.

"Interactive TV, Jack. Wave of the future, man"

— Dennis Hopper in *Speed*

Whither HDTV?

The HDTV research and development effort is based on the mistaken assumption that viewers want higher-resolution programming. Instead, it is well known that what viewers really want is better, more expressive programming. Structured video is part of an effort to enable open-architecture television in which media is delivered and displayed using a scalable and flexible architecture that is not dependent on a specific resolution or aspect ratio standard. In the digital age bits are bits as Nicolas Negroponte would say. If your television is a general purpose digital computer, then it could present to you a myriad of imaging standards, those available today and these to be developed tomorrow.

Structured Video

Structured video provides a method for preserving important information about media objects that enables a wide range of applications, including: more sophisticated compression methods, the ability to manipulating content in semantically meaningful ways, greater flexibility in post production, facilitate the sharing content among multiple projects, and a high level of flexibility in media delivery options, as one media message

could be processed on a wide range of deliver platforms that support a different set of capabilities. Current methods of media production fail to retain much of the information that could make compression encoders and decoders work much better. The application of structured video to compression encoders and decoders is well documented (Bender, et al, 1990), however, it's applications in the cinematic domain is an area of current investigation.

Structured video represents video sequences in terms of component parts

The Media Lab has been investigating alternatives to pixel and frame based temporal image streams. Structured video provides a framework for representing video sequences in terms of component parts which may be rendered and composite as a temporal sequence according to a script (Agamanolis, 1996). These components include two-dimensional and three-dimensional objects. Structured video bit streams are resolution, aspect ratio and frame-rate independent. This approach to encoding images provides dramatically improved compression efficiency, the possibility to play on a variety of devices and the ability to manipulate the data in semantically meaningful ways (Bove & Watlington, 1993).

"with the advent of the [digital] code, the emphasis on perspective returns. Moving-image art can now embrace it in an emphatic way. When the image is a three-dimensional database, perspective becomes a temporal as well as spatial phenomenon. It is a strategy that is intrinsic to the [digital] code. Painters, photographers and filmmakers could not realize the full potential of this desire. But now we can unfold and elaborate that which could only be indicated in earlier media"

— Gene Youngblood (1989)

Structured video enables the development of new story forms

Computer-based media is beginning to compete seriously with video and film as a means of artistic expression. Born of the marriage of traditional video, structured image representation, 3D computer graphics, and cinematic storytelling, structured video enables new approaches to production and facilitates on demand synthesis of images.

Structured video shifts our thinking from frames to scenes

With structured video the frames are assembled from several component objects which may be of different types. For example, our sequence may consist of a 3D model of a set and actors in the form of 2D video objects. For presentation, these elements are composite together.

Traditionally, video has been an analog or digital representation of frames. A succession of frames, typically at 30 frames per second, constitutes a video image. Notions of pixels, scan lines, interlace, and frames are discussed at length in Mathias & Patterson (1985) and Benson & Whitaker (1990) so we will not concern ourselves with those here. The issue we are concerned with is that these frames consist of images without structure, without representation, except for a collection of pixels or variations in voltage. Once we create a video image, it is difficult to make changes to it, except to overlap and com-

"The key to the future of television is to stop thinking about television as television. TV benefits most from thinking of it in terms of bits. Motion pictures, too, are just a special case of data broadcast. Bits are bits."

—Nicholas Negroponte, *Being Digital*

posite it with other images, or to edit collections of frames together. All these operations are carried out without any 'knowledge' of the content of these frames. Much like the notion of 'paint graphics.'

In contrast to traditional video, structured video, is more akin to 'object-based graphics.' A structured video representation consists of various component objects that are combined to make up a video image. This image is more than just an image, therefore it may be more accurately described as a 'video scene.' Structured video objects may consist of 2D pictures or graphics, traditional video objects, text, 3D models, content-specific annotation, production data, etc. A good discussion of the notion of a video objects may be found in Higgins (1994).

Structured video takes away the notion of a specific frame in the production process, however, physical, technological and economic factors influences the form of *projection*, and thus the issue of a frame comes back into the *picture.* Structured video does not make the frame go away entirely, though. You could say that it creates for us a need to *think* about it more than ever, for now the notion of frame becomes a *dynamic* thing. We no longer will 'compose shots' but will 'model shots' and encode them with our intentions.

Point of View and Narrative

Point-of-view is a means of structuring narrative and a powerful mechanism for the manipulation of the viewer's subjective position, revealed through the process of narration. Point-of-view is a very tricky term used in so many different ways. In this thesis, when I refer to point-of-view, I am specifically using the term to "refer less to the character's optical facility than to the narrators process of constructing the spectator's attitude of the film" (Stam, Burgoyne & Flitterman-Lewis, 1992, p. 85).

Edward Branigan (1984) makes a compelling argument that point-of-view is the most significant aspect of narrative structure. Nick Browne (1982) argues that narrative authority and 'spectatorial identification; produces a perspective which overrides the represented, optical point-of-view.

As we design a multiple point of view experience, we have to take into account that a meaningful form of interaction can not be based on optical manipulations alone (e.g. *Aspen Movie Map* and *Museum Movie*). Point of view must be based on a change in subjective positioning which is realized through the carefully crafting of the experience by an author through the

For me the word 'entertainment,' even more than meaning a diversion, means involvement. Sometimes even bafflement, like a magic show. Somehow the commercial cinema, as it's called, has become very frightened of that aspect of cinema. To capture and involve an audience I believe the film-maker of the future will be constructing his film and characters showing many levels and sides to their lives. We will see hidden journeys and, inevitably, by doing so, the film-maker will expose aspects of his own 'secret life'—far more than he does today—and however oblique this might seem in terms of the plot of the film, I'm sure the audience of the future will have some innate sense and understanding of it. I think we are, at present, going though a phase where entertainment is solely equated with distraction. A half-escape.

—Nicolas Roeg

process of narration. The author must consider the subjective and emotional effects of each and every element of the experience, particularly framing, camera position, editing, cause and effect relationships, temporal relationships, etc. For example, Joel and Ethan Coen make extensive use of 'shakycam' and low camera angles to emphasize the outrageous nature of particular characters in their films.

If we think of the process of narration as the "measured plan or pattern of a narrative" (Chatman, 1993) it follows that the axis of time is very important to how a narrative is revealed. Chatman distinguishes between viewer's time and story time. As authors we also take into account the human drive of curiosity, anticipation of what will happen next. If we have to stop and decide on something, as is done with many branching narratives, this takes us out of our reverie! One of the things a good author does is constrain these paths. Freedom to influence, yet still constrained.

Constraints, guidance & subjunctivization

Digital media technologies remove many of the constraints that storytellers have been working with and around for thousands of years. Yet, constraints are a good thing as far as Laurel (1991) is concerned, she reminds us that "constraints are necessary to contain action within the mimetic world." No matter how sophisticated our computer, no matter how many terrabytes of data we have at out fingertips, our computer can't present all possible worlds simultaneously.

Bruner (1986) discusses the "performance of meaning under the guidance of the text." Our narration process must make it possible for the viewer to 'write' their own virtual 'text.' Bruner discusses three essential elements of the process: the triggering of presupposition, subjectification, and multiple perspective. Through the *triggering of presupposition* we create implicit rather than explicit meanings, expanding the viewer's interpretive options. *Subjectification* is the depiction of reality through the 'filter' of a character's consciousness, rather than an omniscient, 'third-person' eye. Through *multiple perspective* we see the world through a set of 'views' each which captures a different part of the world.

These three elements, when combined, succeed in what Bruner calls *subjuntivizing reality.* To be in the subjunctive mode is "to be trafficking in human possibilities rather that in settled certainties" (Bruner, 1986, p. 25-26). This *subjunctivization* is the key to the expressive power of narrative and its ability to mean more than is said. It also offers some insights into why a purely 'first-person' experiences is less engaging

than a mixture of 'first-person' and 'third-person' views.

The problem of person in narrative

Creating first person narrative experiences is much tricker than you might at first think. There is a temptation to confuse multiple optical perspectives with multiple perspectives in the narrational sense. Let's take a look at two examples, one from cinema, and the other from interactive media, as there is much to learn from these.

Lady in the Lake

Lady in the Lake (Robert Montgomery, 1946) is a film that was shot almost entirely as an elaborate point-of-view shot from the perspective of detective Phillip Marlowe, who acts as a diegetic narrator. Marlowe is investigating a crime and as he goes about his business we see everything from his vantage point, we see what he sees, we hear what he hears. Director Robert Montgomery went through great lengths to create the first person experience, including the reflection of Marlowe in mirrors and making the movements appear to match Marlowe's movements, including bumping into things.

Voyeur

A more recent example is *Voyeur* (Philips, 1994), a CD-I title that also attempts to provide a first-person experience, however, new technology allows the experience to be interactive. The game situates you in an apartment from which you are able to look out of your window. The story is about Reed Hawke, a U.S. presidential candidate whose apartment you can see from your own. Hawke tries to suppress a dark family secret while you videotape scenes of lust, betrayal and revenge. Your goal is to get this information police in order to destroy Hawke before he can destroy you.

Critique of Lady in the Lake and Voyeur

Lady in the Lake and *Voyeur* fail to provide a satisfying, emotionally engaging experience. In its day audiences failed to identify with Marlowe's character in *Lady in the Lake* and *Voyeur* fails to get you into the 'head of the character' making the same mistake as *Lady in the Lake*. These experiences serve as a reminder that subjective perspective is most effectively constructed out of a combination of shots, each with a different perspective. Both of these examples fail to make effective use of the three essential elements of the narration process: the triggering of presupposition, subjectification, and multiple perspective. To 'see' the entire story from a character's optical perspective actually ends up reducing the subjectivity of the viewer's experience. First person in the cinema and related

visual mediums is more complex than one would think. Some theorists have shed additional light on this issue.

Creating subjectivity

Nick Browne (1980, 1982) makes a compelling argument that narrative authority and 'spectator identification' produces a perspective which overrides the represented, optical point of view. Shifting subjective positioning in a story requires more than just a change in optical perspective. Branigan (1992) reminds us that "representing true subjectivity depends upon exploiting a spectator's shifts of attention and memory through a blend of subjective and objective techniques," and Brinton (1947) writes "striking cinematic effects depend on what the spectator apprehends outside the scope of the camera" and goes on to suggest that narrative is not defined by specific material forms but rather is defined through the spectator's predisposition to make narrative and meaning using whatever is at hand.

We experiment with the concept of controlling narration and subjective positioning and implement some of these notions in the design of *Two Viewpoints,* described in the next chapter.

Little Dorrit

Little Dorrit (Christine Edzard, 1988) provides an excellent examples of a multiple point of view narrative. Like the Dickens novel, the film presents a rich and panoramic story set in 18th century London. But Edzard has added a new Twist: *Little Dorrit* is two feature films that present substantially the same narrative yet each is narrated from a different perspective.

The first part, *Nobody's Fault,* is narrated from the perspective of Arthur Clennam, while the second part, *Little Dorrit's Story,* is narrated from the perspective of Little Dorrit. Together, the two films are approximately six hours long. Neither film is complete on it's own, the two parts are highly interdependent. *Nobody's Fault* does not conclude the story, nor does *Little Dorrit's Story* provide sufficient information to explain everything that is happening in the story. The scenes which are similar in the two films vary in details of performance, dialog, settings, costumes, etc. While you are watching *Little Dorrit's Story* you realize that what you saw in *Nobody's Fault* was highly subjective, and challenges your assumptions that an omniscient narrator was at work.

Director and writer Christine Edzard has effectively brought to the screen much of the subjectivity and richness which we find in novels but is often lost when they are translated to film,

Little Dorrit's Story

Little Dorrit is a story set in 18th century London about Little Dorrit, a young woman born and raised in Marshalsea, a debtors' prison. Her father has been imprisoned there for over twenty years, and has become known as "The Father of Marshalsea." The father has no real hope that he will ever be released, however Little Dorrit helps supports her father and domestic tranquility is maintained.

Everything changes one day when Arthur Clennam, a middle-aged gentleman recently returned from a long trip abroad, takes an interest in Little Dorrit and helps the Dorrit family financially. This plot is only a framework from which a myriad of subplots and complexities are intertwined in fascinating ways.

The phrase "Nobody's Fault" represents society's collective guilt and lack of desire to remedy a complex problem. *Little Dorrit* is a masterful mix of entertainment and social commentary offering the viewer an opportunity to experience the story as two separate films, each with a different point of view, the first, Nobody's Fault, is told from the perspective of a privileged person in society, while Little Dorrit's Story is told from the perspective of someone who suffers under social institutions.

and she does this with an ideological savvy and attention to story detail often lacking in cinematic adaptations from books. The film's unique structure makes it possible to experience a subtle and complex social comment from two different viewpoints, the first that of a privileged person, who can easily say it's nobody's fault, while the second one has you empathizing with someone who directly suffers from the social institutions of society, someone who can clearly see that the fault must lie somewhere.

Little Dorrit also sheds some light in the limitation of providing a multiple point of view narrative with two linear films, it is evident that Edzard had a tough time deciding between repeating too much of the material from the first half and not showing enough of it to make the differences in the points of view clear. This is where an interactive, multivariant presentation would be a big win.

I can't think of a film that has ever close as close as *Little Dorrit* has to replicating the richness and dimensionality of a Charles Dickens novel. Their sheer length alone makes adaptation to the screen within the confines of a standard 120 page script close to impossible, let's face it, there are many stories that simply don't fit into the format. Edzard was able to write and direct *Little Dorrit* without diluting much of the story or it's socially conscious message for sake of narrative efficiency, clarity or running time. *Little Dorrit* is quite an ambitious film and will hopefully inspire us to achieve the same artistic level with our new interactive and multivariant media forms. *Little Dorrit* is a rare example of a long form multiple point of view narrative and provides a valuable opportunity to compare and contrast multiple narrations of the same story.

He Said, She Said

Another film that attempts to narrate the same story from two different perspective is *He Said, She Said* (Ken Kwapis and Marisa Silver, 1991). Dan and Lorie are two journalists working in the same office. More often than not they have opposing view of the issue in question. Deciding that this is has potential as a television show, a producer gives them their own program where they can give their opposing views on various issues. As they work together and get to know each other, the events that occur in their lives are replayed in the film from two perspectives, his and hers. *He Said, She Said* is a much less ambitious example than *Little Dorrit,* nevertheless, it's interesting enough and at times entertaining.

The Cinematographer's Activity

The screen/frame provides a field for aesthetic expression. It provides the moviemaker with a way of creating psychological closure, clarifying and intensifying the viewer's experience. During preproduction the director and the cinematographer attempt to work out the general flow of the shots and work out the coverage needed for scenes. By coverage I am referring to additional shots which overlap the action captured in the master shot, but from a new angle and framing.

The *shot* is a uniquely cinematic construction. The structural aspects we consider when we construct a frame include: 1) the aspect ratio, 2) object size (comparison, reference, knowledge of objects, relation to screen size, context and reference to people), and 3) image size (size constancy, image size vs. relative

energy, spectacle). In the following discussion we will concern ourselves primarily with the issue of aspect ratio and the differences inherent in small vs. large display screens.

Aspect Ratio

Before the advent of wide screen movies in 1953 1.33:1 was the dominant ratio for movies. Much lore surrounds how this number came to be, but for all practical purposes it was pretty much arbitrary. With the advent of wide screen came a proliferation of aspect ratios. 16mm film and later television adopted the 1.33:1 aspect ratio. After the dust settled on the proliferation of aspect ratios during the advent of wide screen in the 50s, 1.85:1 became the theatrical standard in the United States and 1.66:1 is common in Europe. Contemporary anamorphic wide screen is 2.35:1. For an excellent discussion of aspect ratio, I suggest Schubin (1995).

Cinematographers have over the years taken the issue of aspect ratio into their own hands, for example *Intolerance, Chinatown, Natural Born Killers, A Short Film About Killing,* which use elements of the mise en scène or graduated filters on the lens to visually alter the effective aspect ratio. Why not make this a dynamic media element instead of an industry standard? Just because a screen has a particular aspect ratio does not mean you have to use all of the screen real estate. For example, *Theremin* was recently released theatrically as a 1.33:1 movie within a standard 1.85:1 theatrical release print.

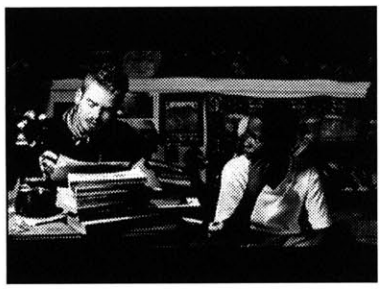
Perceptually, we may use either an inductive or deductive approach when making sense of an image depending on the size of the image

A large screen or immersive display lends itself to a deductive information processing approach on the part of the viewer, while a small screen lends itself to a more inductive approach (Zettl, 1990). One creative opportunity made possible by structured video is the presentation of a dynamically composed frame depending on the particular viewing device being used. When we view something on a large screen we tend to move from the general to specific details. On the other hand, with a small screen we tend to move from the details to the general. This is due to an interplay of what we can see in the frame as well as the editor's decision of how to cut a sequence. Editors recognize the fact that the smaller television screen lacks the 'aesthetic energy' of the large cinema screen, therefore they tend to favor more close-ups and insert shots. Video editors have a tendency to build a sense of a scene from several shots rather than a large panoramic view.

And this makes sense, since panoramic, wide shots, which have tremendous aesthetic energy on a large screen lose their impact on a small screen. Thus, when you letterbox a film that was composed for a large screen, the wide, panoramic shots lose their energy. Ideally we would replace these with a sequence of shots that build the sense of the scene inductively rather than deductively. The figure on the next page shows a wide screen composition from *Slack Trek* (Julie Chang & Joe Rubin, 1995), a feature film I worked on as cinematographer. We shot the film in Super 16mm which provides a 1.66:1 aspect ratio. I composed scenes for the standard 35mm theatrical release print ratio of 1.85:1 with the primary objective to

Commandant teaching pocket gonzo, "deductive" frame.

From *Slack Trek*, courtesy of 21st Century Fish Productions, used with permission.



Commandant teaching pocket gonzo, "inductive" sequence.

From *Slack Trek*, courtesy of 21st Century Fish Productions, used with permission.

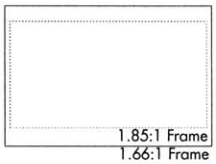
compose scenes for large screen viewing. In this particular scene the rebel leader "Commandant" is teaching political philosophy to Pocket Gonzo. I originally composed this shot to be 'read' deductively, inviting the viewer to look around and deductively perceive what's going on in the scene. For example, the titles of the books on the table reveal Comandant's political orientation. The first frame on the left represents a smaller version of this scene which has been letterboxed to preserve the original aspect ratio. Reduced and cropped for a small display the original composition is preserved at the cost of not being able to recognize the titles of the books and the character's reactions. Restructuring the scene as a series of shots changes the composition considerably, reinterpreting the scene in constructive cinema language rather than master scene cinema language, however, the titles of the books and recognition of the character's reactions is preserved. In this case I would argue that narrative coherence is considered more important than a specific aspect ratio or composition. For the smaller display, the inductive approach proves more effective.

Cinematographers today often have to compose for multiple aspect ratios

Since the beginning of the home video revolution directors and cinematographers have had to face the issue of composing for multiple aspect ratios, since a majority of their audience will see their films on video. There are basically three approaches to this problem: (1) Ignore the 1.33:1 video frame when composing for 1.85:1 or 2.35:1 wide screen, (2) compromise the wide screen composition in order to have a pleasing composition for the 1.33: 1 frame, or (3) compose primarily for the 1.85:1 or 2.35:1 frame, giving due consideration to how the scene will play out on a 1.33:1 aspect ratio, this approach may involve the use of alternative formats (e.g. Super 35mm) in order to facilitate this multiformat approach.

Never Met Picasso (Stephen Kijak, 1996) is a feature film I shot in Super 16 (1.66:1). I chose to compose the film primarily for the 1.85:1 aspect ratio as the film will be optically 'blown-up' to 35mm for theatrical exhibition. I wanted to avoid sacrificing the composition in 1.85:1 for the 1.33:1 frame. This did yield optimal framings throughout the film for the theatrical release, however, this will lead to some unusual compositions for the 1.33:1 television frame, requiring panning and scanning during the video transfer process. Ideally I would like to transfer it to video 'letterboxed' as a 1.85:1 film, however, to this day distributors are hesitant to release letterboxed films except as special editions on laserdisc for the film buff crowd. Optimizing the film for 1.85:1 was a decision that the director and I, as cinematographer, made in advance. In some situa-

Original 1.85:1 aspect ratio frame



Super 16 Frame

1.33:1 aspect ratio frame with pan

Andrew (Alexis Arquette) and Lucy (Georgia Ragsdale) in *Never Met Picasso*, courtesy of Mighty Reel Productions, used with permission.

tions this strategy required me to try to visualize how the scene would play out in a pan and scan scenario to avoid a total mess later on down the production pipeline. For example, the scene with Andrew and Lucy (shown below) plays better as a two shot, however, for video it will not only require a pan, but a little bit of cropping as well to avoid seeing part of Andrew in the final frame, an aesthetic choice. One thing that works in our favor is that the 1.66:1 aspect ratio provides a little bit of additional area at the top and bottom of the frame which I can make use of in the video transfer. For this reason I had to 'protect' 1.66:1 which meant that even though I was composing the film for the 1.85:1 frame line, I had to make sure there were no extraneous objects (e.g. lights, boom mike, etc.) in this area of the frame because it might appear in the video release.

The 35mm frame gives you alot more to play with in this area, since the 35mm frame is 1.33:1 and 1.85:1 standard wide screen movies are shot with the standard 1.33:1 frame, sometimes a hard matte is used in the camera to block off the extra

area, and sometimes the film is actually shot 1.33:1 which provides some additional play when it comes time to do a video transfer of a theatrical film.

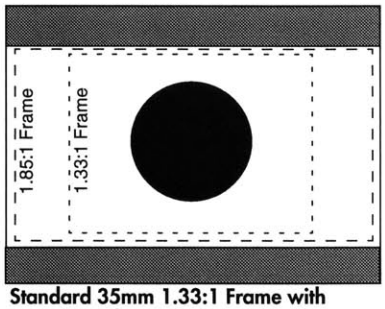
As a cinematographer, I would like to have an alternative to this either-or situation. Why can't composition be adynamic process? Our moviemaking tools should allow us to optimize the framing for different presentation media and is one of the issues explored with *Two Viewpoints* and the *SAR Toolkit.*

Blade Runner and T2 have been transferred to video with different results

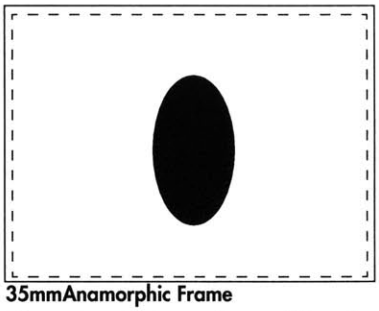
Blade Runner (Ridley Scott, 1982) and *Terminator 2: Judgement Day* (James Cameron, 1991) provide two contrasting examples of wide screen films that have been transferred to video with very different results due to considerations during the preproduction and production stages.

Jordan Cronenweth shot *Blade Runner* before home video became the large market it has become by the time Adam Greenberg shot *Terminator 2.* The compositions in Blade Runner make use of the wide screen very effectively, however, the film suffers alot in the transfer to 1.33:1 since much valuable screen information is lost. Imagine reading *War and Peace* with every third word removed. Many contemporary films take another approach, they compose primarily for 1.33:1 and the extra screen area is just that, 'extra.' This is commonly referred to as 'protecting 1.85:1,' a pretty sad state of affairs. The strategies for composing for 1.33:1 and 1.85:1 are sometimes at odds in my opinion. This approach of shooting basically a 1.33:1 frame and protecting the 1.85:1 edges fails to exploit a powerful elements of cinematic expression: dramatic framing.

On the other hand, James Cameron and Adam Greenberg composed *Terminator 2* primarily for the 2.35:1 wide screen aspect ratio, however, they also thought about how the frame would survive the transfer to 1.33:1 and decided to use a strategy that would enable them to optimize the framing so that they could accomplish both goals. Instead of shooting 2.35:1 anamorphic, Greenberg opted for the Super 35mm format. Super 35mm provides the largest possible negative during shooting by using the area reserved for the optical sound track as additional image area. This format provides for better framing options when transferring the film to video with a 1.33:1 ratio. It does, on the other hand present some technical sacrifices in that prints made from it do not look as good as prints made from an anamorphic negative. This is because the Super 35mm negative has to be optically printed to create either 1.85:1 or 2.35:1 anamorphic theatrical prints. But that is neither here nor there as far as the current discussion goes.

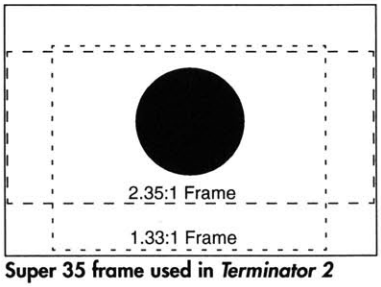


Standard 35mm 1.33:1 Frame with 1.85:1 Hard Matte



35mmAnamorphic Frame

The 2.35:1 ratio picture is "squeezed" into the 1.33:1 frame by an anamorphic lens and is then 'unsquozen' for projection.



Super 35 frame used in *Terminator 2*

Rent some videos, make some popcorn and take a look for yourself

Blade Runner is available on video in both the 1.33:1 and 2.35:1 letterboxed versions as is *Terminator 2* which was released on laserdisc in both the 1.33:1 and 2.35:1 letterboxed versions.

These two films provide an excellent opportunity to compare how the reframing for 1.33:1 was done on a scene by scene basis, and an opportunity to compare the effectiveness of the different shooting and reframing techniques.

Terminator 2 was shot using Super 35mm with the primary frame being a 2.35:1 'window' sharing a common top of frame with the 1.33:1 'window' thus providing more room for maneuvering during the pan and scan process. *Terminator 2* on a small television looks much better than most other wide screen films due to this approach which avoids the problem of close up shots looking too tight in the 1.33:1 version, an example of this is the close-up Leon in the interrogation scene in *Blade Runner.*

Master Scene Cinema Language

D. W. Griffith is credited with developing what has become known as the Master Scene Cinema Language which during the 1940s evolved into the seamless and efficient 'language' of the Hollywood film. There are other cinema languages, but for the sake of simplicity I've chosen to limit the discussion and use of cinematic language in this thesis to Master Scene Cinema Language. For a discussion of interpersonal, constructive, and other cinema languages, see Richards (1992).

"The close-up may sometimes give the impression of a mere naturalist preoccupation with detail. But good close-ups radiate a tender attitude in the contemplation of hidden things, a delicate solicitude, a gentle bending over the intimacies of life-in-the-miniature, a warm sensibility. Good close-ups are lyrical; it is the heart, not the eye, that has perceived them"

— Béla Balázs, *Theory of the Film*

A cinema language serves as a blueprint for communicating ideas using the cinema as a medium of expression. The shot flow strategy in Master Scene Cinema Language involves starting with an establishing shot that 'covers' the entire scene in a single shot, typically wide. The framing usually covers all of the relevant scene elements and characters. Coverage of the scene is made up of one or more additional 'setups' typically photographed with an establishing shot, a medium shot and a close-up. Typically only portions of the scene are recorded at successively tighter framings. This coverage allows the director and editor more flexibility during the editing process, allowing cuts to be made in accordance to the emotional ebb and flow of the scene and the actors performances, in addition to providing flexibility in the manner that both characters and space are revealed.

Along with the actor's performances, the blocking of action and composition of the frame are at the center of the filmmaking process. From a physical standpoint, this is the manner in which, along with editing, the director carefully controls the viewer's subjective position in a scene.

Master Scene Language in practice

Feature films are typically shot using master scene cinema language. This implies that we shoot a scene from multiple camera positions, with significant overlap in the shots between camera setups. By shooting this scene from different camera positions, each with it's unique framing and composition, the

editor has several option of when to cut and can control rhythm, subjectivity and identification.

In preproduction the director and I, as the cinematographer, discuss the script, especially in terms of characters and locations, and how the specific locations reflect the characters. Rough action is then sketched out with storyboards and blocking diagrams. But these are only a start. Once we get on location, we usually refine our shot list based on considerations regarding the location itself. Then we typically have a blocking rehearsal with the actors. During the blocking rehearsal I watch the action and walk around the set, considering how to best and most efficiently cover the scene. The amazing thing is, no matter how clearly the director and I imagined what the shot will look like in preproduction, once we start working with the actors, our preliminary notions usually change to some degree. The actors, through their performance on the set, often show us new ways to cover the scene.

In effect, the master is for the most part a dress rehearsal, it's a shot we want to get out of as quickly as possible, especially in interpersonal scenes. Actors will usually save their best performances for the medium and close shots. All this coverage may seem like a waste, but on many occasions it made a difficult scene easier to edit, and on two occasions when there were technical problems with the film, the editor was able to 'cut around the problems'.

Alfred Hitchcock's films are notable exception to the language of Master Scene Cinema Language. See Truffaut (1984) for an excellent discussions of Hitchcock's 'constructive' approach to filmmaking.

3 Two Viewpoints

Two Viewpoints is an interactive, multivariant movie currently running on the Cheops image processing platform. We produced *Two Viewpoints* in order to demonstrate and experiment with the concepts of dynamic framing and sequencing in real-time. The movie is multivariant in terms of subjective perspective as well as frame size. The viewer interacts with the story in real time by choosing Kathy's perspective or John's perspective, our two protagonists. *Two Viewpoints* takes advantage of structured video representation and the abilities of the Cheops imaging system to present a movie that is assembled from a collection of media objects in real time according to a script. In this section we review our objectives and describe the process of design, production, postproduction, and presentation.

Objectives

The design, production and delivery of *Two Viewpoints* was driven by various technical and aesthetic objectives.

Test the theory regarding interactive narrative & narration

The narrative and film theory cited in the previous section provides a foundation for thinking about the creation of an interactive experience in which the viewer is given the opportunity to influence the narration of the story; however, theory must be balanced with production and experimentation. *Two Viewpoints* should offer some feedback on the viability of the theoretical foundation discussed in the previous section.

Explore production techniques for structured video and Cheops

We wrote the story and designed the scene independent of a specific shot sequence, aspect ratio, framing, and screen size. Framing and editing decisions, and the specific images required to convey the story, were encoded in the form of an *Isis* script (Agamanolis, 1996). This script, when executed by the Cheops imaging system (Bove & Watlington, 1995), responded to user interaction and rendered the scene in real time.

The use of structured video representation in general, and the

Not the Royal We

Two Viewpoints was a highly collaborative project, the key players include Shawn Becker, Stefan Agamanolis, Araz Inguilizian, Tony Romain and myself under the direction of Professors Glorianna Davenport and Michael Bove at the MIT Media Laboratory.

Stefan implemented the interactive playout of *Two Viewpoints* on the Cheops imaging system using *Isis*, the scripting language he developed which is described in Agamanolis (1996) and was a collaborator during every phase of the production.

Shawn Becker created the 3D model of the set from 2D photographs using tools he developed and was responsible for the overall production design. His methods for creating 3D models from uncalibrated 2D views is described in Becker & Bove (1995).

Araz Inguilizian designed and implemented the structured sound for *Two Viewpoints* using his tools described in Inguilizian (1995).

Tony Romain wrote the script for *Two Viewpoints*, collaborated on the interactive design, and directed the production.

I served as producer of *Two Viewpoints* and developed the story and interactive design in collaboration with Tony Romain and David Kung. I worked with Shawn, Stefan, and Araz throughout the post-production phases and developed the *SAR Toolkit*, described in the next chapter.

Several people provided valuable assistance during the phases of design, production, post-production and implementation. Their names may be found in the acknowledgments.

Cheops imaging system in particular, enable us to deliver a movie which is assembled in real time from a collection of media objects. This mode of content delivery makes it possible to deliver a movie that is framed and sequenced dynamically at the time of delivery with the possibility of viewer interaction.

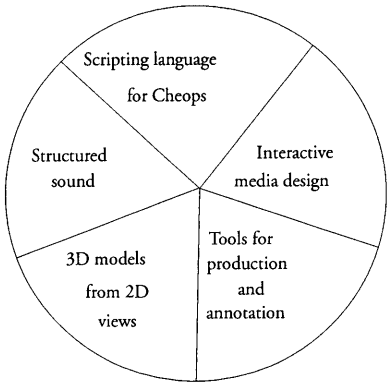
Two Viewpoints provides a test bed for building tools to facilitate the production of structured video projects. We inform our tool design through the process of production and delivery.

Viewer interacts with the *process*, rather than the *structure*, of narrative

We are interested in designing interactive experiences based on the process oriented theory of narrative and narration set forth by Branigan (1994) and to demonstrate that this theory of narrative, while deconstructionist, sheds light on some salient features of narrative construction that are directly applicable to the design of an effective interactive narrative experience. An approach for interactive narrative that has not been explored much is giving the viewer control over aspects of the *narration* (process) of the story rather than coarse control over the *narrative* (structure). Coarse control over the narrative is all too common with branch-structured or collage-structured interactive narratives.

To make an apple pie from scratch, you must first invent the universe.

— Carl Sagan



Two Viewpoints is a multifaceted research project comprised of several slices of *pie* (personalized interactive entertainment modules) so grab a cup of coffee, sit back, relax, and enjoy the show...

Let the viewer choose a subjective position within the story

Two Viewpoints should give the viewer the ability to vary the manner in which the movie is narrated. In this case the variability in narration gives the viewer the option of seeing the movie from the perspective of either character, or both, in a manner which combines different perspectives into an edited sequence.

Branigan (1994) reminds us that point of view is an intricate and tightly crafted construct. Therefore, the moviemaker must carefully design the experience in order to give the viewer subtle yet significant control over the process of *narration* rather that the structure of *narrative*. Construction of narrative is a more complex process, and places high demands on the viewer if it is to be meaningful. Typical branch-structures are not interesting from a narrative perspective. Therefore, with *Two Viewpoints*, we focus on the issue of narration and the potential for interactive narration.

The Viewers Experience

Two Viewpoints is a short narrative sequence structured such that the viewer's empathy can lie with either the character of John or the character of Kathy. The viewer is given three ways

of interacting with the story: (1) change the optical position of the camera, (2) change their 'subjective position,' or (3) change the frame size; all of these interactions are supported in real-time while the sequence plays.

Not just sequences anymore...

Previous work in video orchestration has dealt primarily with the sequencing of shots, providing the moviemaker with control over montage, but not mise en scène. With *Two Viewpoints* we explore one of the greatest advantages of structured video representation: control of both montage *and* mise en scène, allowing the moviemaker to structure the narration. Video orchestration systems have been most successful with documentaries and catalogs, however, narrative requires a more tightly authored environment. Giving the moviemaker control over mise en scène as well as montage makes for a more complete authoring environment.

The evolution of *Two Viewpoints*

In the spring of 1995 I collaborated as cinematographer on *The Yellow Wallpaper* with Tony Romain, who wrote the screen play and directed it as his Boston University thesis film. The conflict in the story, and the radically different perspectives of the characters became a starting point for *Two Viewpoints.* During pre-production meetings with Tony and I discussed scenes from Kathy and John's perspective, and eventually a scene that takes place between John and Kathy stood out as a short, self-contained, sequence that would tell a good story within the constraints imposed by Cheops. The sequences had a strong and immediately recognizable conflict between the characters, providing viewers with two clearly different perspectives.

Adapting the story

Story development is never an easy, straightforward process. Some elements of the original scenario were going to either take too long to implement or simply made the story too complicated. This sequence had to play out in less that 120 seconds and get it's basic point across. Tony collaborated with us to develop a short, multivariant sequence that would demonstrate the creative potential of structured video and at the same time present an evocative, dynamic story.

In order to clearly communicate the two threads in the story, Tony wrote two scripts, one from Kathy's perspective, the other from John's. In preproduction we combined the shot

Gilman's *The Yellow Wallpaper*

Charlotte Perkins Gilman's *The Yellow Wallpaper* was published in *New England Magazine* in 1892. The story concerns a female narrator's descent into madness. Gilman wrote the story based on her own painful experiences as a critical commentary on the conflict between women's familial duties and public responsibility. Gillman wrote this 'fictional exorcism' in hopes of bringing about change in the methods used to treat 'nervous prostration,' in fact, she names Dr. S. Weir Mitchell in the story, under whom she underwent the treatment that she wrote about (Erskin & Richards, 1993).

Romain's Adaptation

Tony Romain's short film is based on Gilman's *The Yellow Wallpaper*, however, it features a new structure:. rather than telling the story through a single narrator situated in the late 19th century, he creates two characters: Kathy is a in a contemporary setting writing a story, the story she's writing is about Kathryn, in the 19th century slowly going insane living in a large colonial estate, The film intercuts scenes of Kathy and Kathryn. The plot of the Kathryn's segments is similar to Gilman's original story, however, Kathy is quite different than Kathryn. Kathy has to deal with an abusive husband, and her escape is writing fiction. The two women become aware of each other. In the film Kathryn is given the option to transcend her situation, Kathy is able to rewrite the story. Romain's approach to the story is interesting: rather than endow a 19th century character with 20th century sensibilities, Romain gives the 20th century woman the "power to rewrite the relationship between her and her society."

lists into one. The changes in point of view are articulated primarily through camera framing and cutting. Two scripts, one shot list, three perspectives. The power and versatility of cinematic language.

The first problem we ran into is how do we script an interactive movie? In our case we wrote two scripts and combined the shot lists. In postproduction we worked out the actual sequences and the manner in which they would play out. Our approach was much the same as the approach one takes shooting coverage for a scene using Master Scene Cinema Language.

It turned out, after considering alternatives, that for a interpersonal scenario like *Two Viewpoints* that it was simplest to read and write two separate scripts in traditional script format which presents the 'essential' dialog, screen direction, and action and deal with the multivariant aspects of the story in the editing phase, especially since we set out to present basically one story, with the variation in the storytelling.

Production

In it's final form, *Two Viewpoints* is a structured video sequence consisting of the following components:

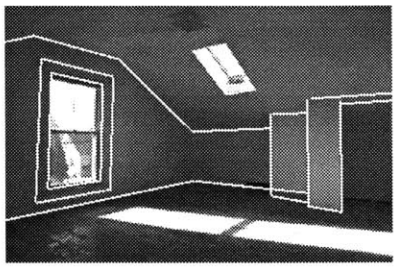
- video objects of the actors annotated with three-dimensional location data and timecode,
- props consisting of video objects annotated with three-dimensional location data,
- a three dimensional model of the scene rendered from two-dimensional photographs, and
- an **Isis** script describing the scene and interactions.

The Set

For the set we used a three-dimensional model of a small attic bedroom. The mode was built from a series of photographs of an attic bedroom using model based video techniques described in Becker & Bove (1995).

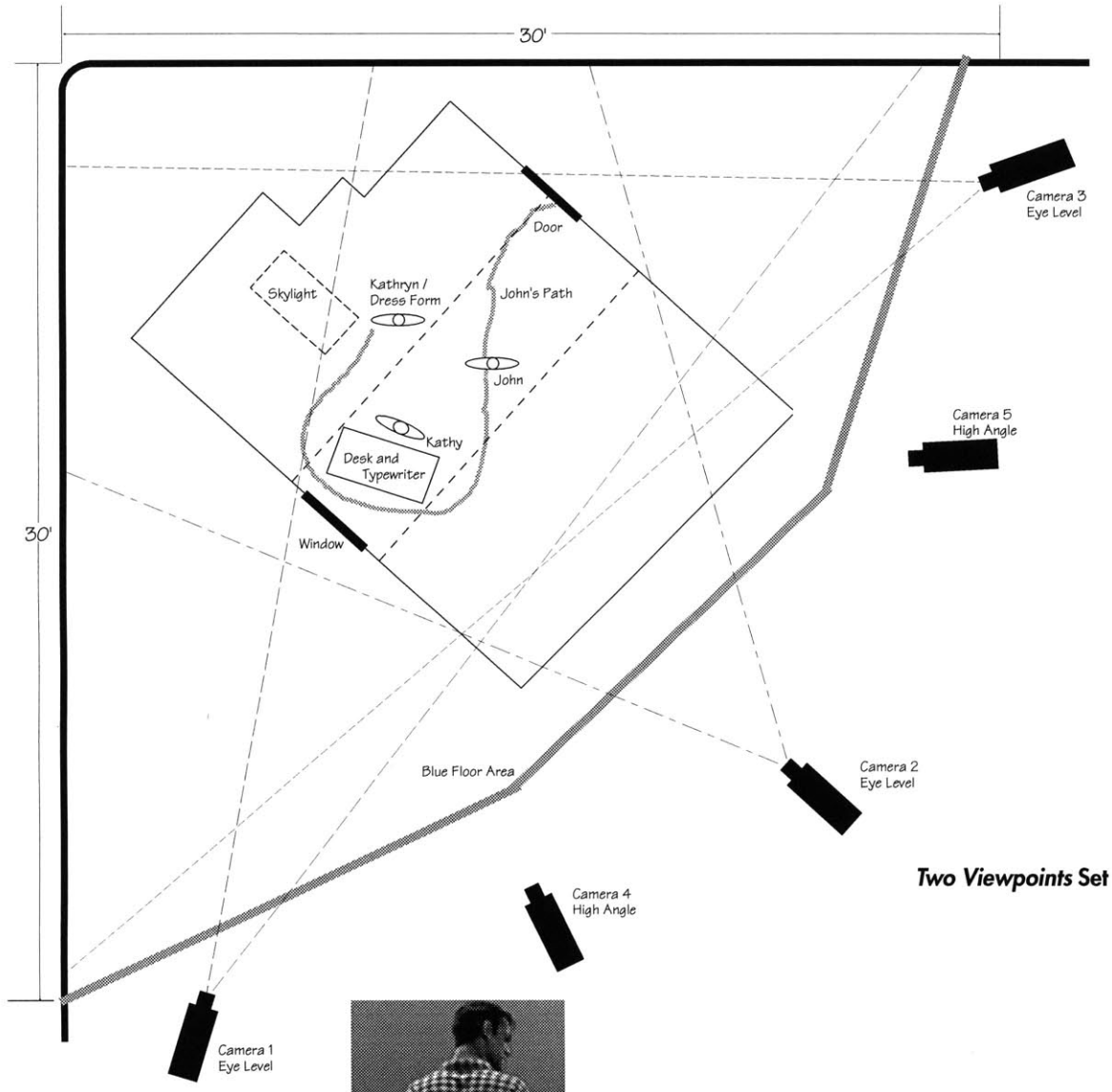
Capturing the Actors

The actors were shot against blue screen in order to provide flexibility in the final framing and blocking of the scene.Using the dimensions of the room from the three-dimensional model, we marked off the equivalent area in the blue screen studio in order for the actors to have a better sense of the space and keep the action within the constraints of the three-dimensional set model.



Model Based Scenes

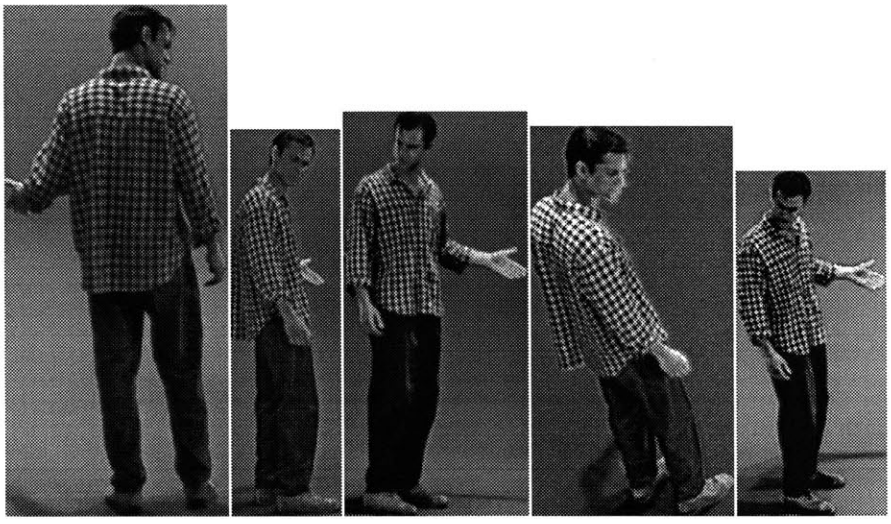
Becker (1995) describes a approach for reconstructing 3D structure and texture from a scene composed of straight edges using the image data obtained from one or more partially overlapping uncalibrated 2D views. Knowledge of geometric relationships among correlated features allows accurate recovery of scene structure and intrinsic camera parameters. Model-based representations of video offers ultra-low bandwidths by exploiting the redundancy of information contained in sequences which visualize what is for the most part a static 3D world.



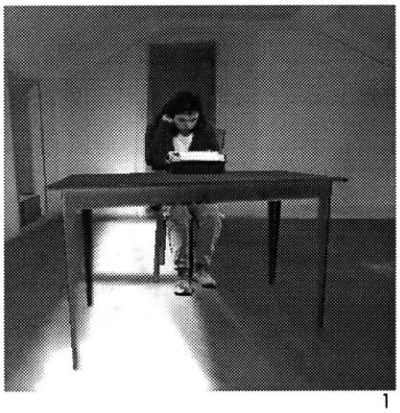
***Two Viewpoints* Set Plan**

John as seen from each camera

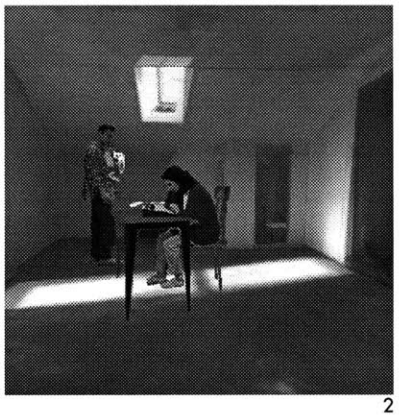
Each frame of John consists of five camera views. Depending on the virtual camera position, Cheops will choose the video object that most closely matches to the required view. Video objects, while less flexible that synthetic models, offers a degree of realism that is high relative to the computational requirements.



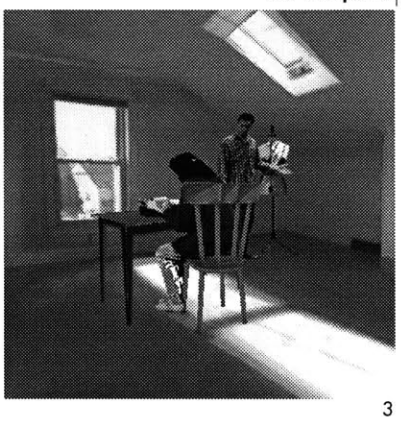
Camera 1 **Camera 2** **Camera 3** **Camera 4** **Camera 5**



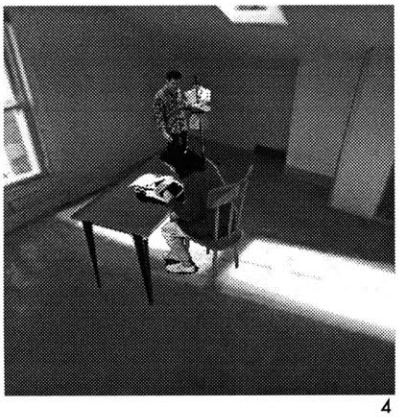
1



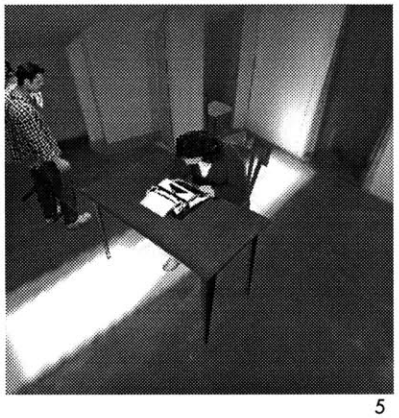
2



3



4



5

Camera Views

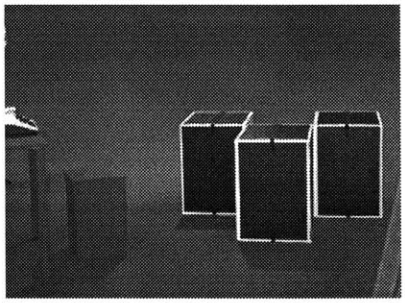
The scene was shot using five cameras, three at ground level and two at a high angle. The high angle cameras facilitated computation of the character positions in the analysis phase.

We used five Betacam SP cameras, three were placed at eye level, two at a high angle. The high camera angles were added in order to facilitate calculating positions of the actors in three-dimensional space using in-house image analysis tools (Becker &Bove, 1995).

All five cameras were synchronized with the same timecode. This facilitated the construction of the actual video objects since for each frame of the actors video object is composed of five video elements, one from each camera view. The starting timecode was added to each of the rgba DAT files with the **XmDat** utility.

Calibration

The master shots required a calibration step. At first we did these before shooting the action, however, we found this to be restrictive. This did not allow the director and cameraperson to change the shot if something in the performance changed or if the actor found it easier to hit a different mark, etc. After two setups, we decided to change our procedure to shoot the scenes with the actors first, and then after we had a circled take (a 'keeper') we would lock down the camera position and focal



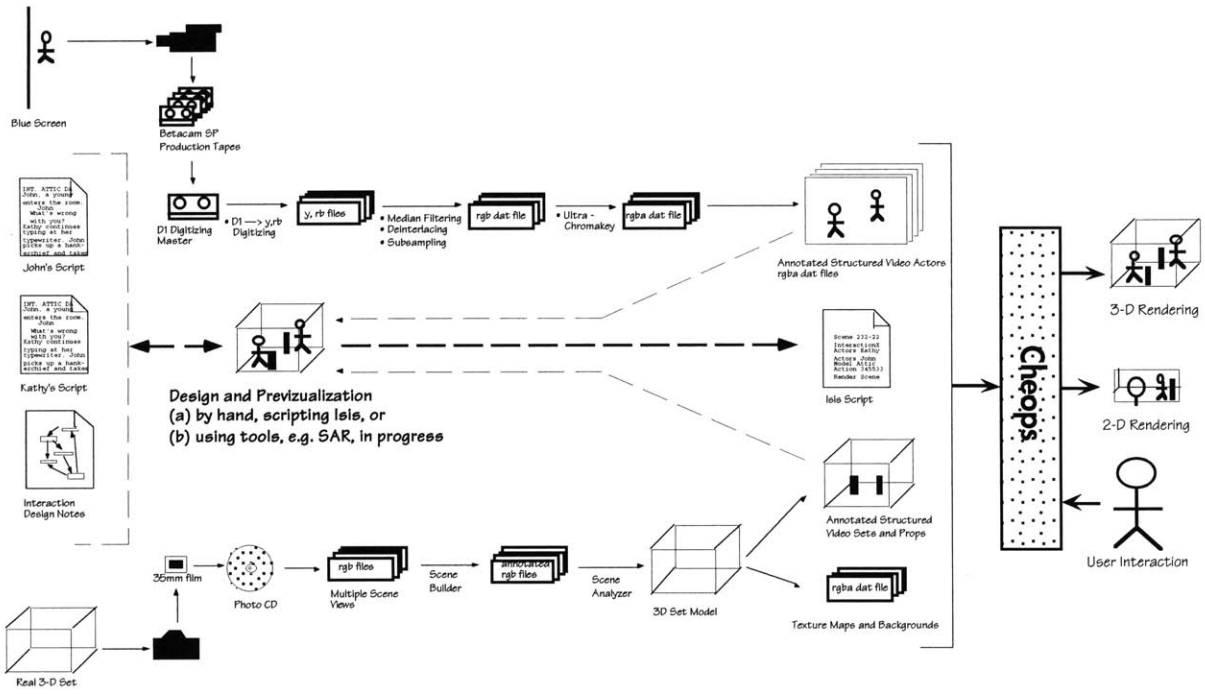
Calibration Cubes

In order for Shawn Becker's analysis software to calculate the 3D positions for the actors, we needed to use 'calibrated cameras'. This required shooting 'calibration cubes' after each setup. This process helps automate the task positioning the actors in the 3D set during playout. The analysis techniques are described in Becker (1995) and Becker and Bove (1995).

length and run the calibration procedure. Close-ups were uncalibrated as these were shot from what we considered in advance to be optimal angles. It was determined that it would be alot of work to match the calibration of the master shot with each of the close-ups and in the end, not productive.

Production Team

Our production required a production team similar to the team that makes up a contemporary video production with the addition of a 'technical team' responsible for (1) the design of the 3D model, (2) design of the interaction, and (3) design of the rendering and playout. This differs significantly from a traditional film crew, and the relationship between the production designer, editor, director lighting director, and the technical team.

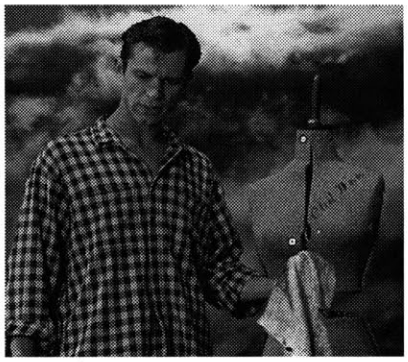


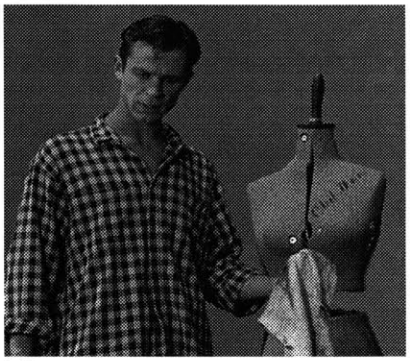
Postproduction

Video objects

Cheops does not currently have the ability to render three-dimensional textured objects in real time. In addition, current techniques for modelling and rendering actors is extremely time consuming and processor intensive and do not yield satisfactory results if realistic representation is desired. Realistic images of natural objects, particularly people. Modelling real locations and compositing actors simplifies shooting logistics


Kathy


Neutral


John

and provides a framework for flexible rendering\

Digitizing Video, chromakey and segmentation

The details of digitizing, chromakey, image processing, segmentation, and assembly of the structured video components is discussed in Agamanolis (1996).

Orchestration

The structured sound and video elements of *Two Viewpoints* are orchestrated on Cheops under the control of an **Isis** script which was written by Stefan Agamanolis.

The Viewer's Experience

Two Viewpoints plays out differently based on your selection using three knobs. You may adjust these at any time during the playout. The first knob allows you to change the spatial position within the master shot. The second knob allows you to vary the frame size. The third knob changes your subjective position within the story, allowing you to see the story along a continuum from either Kathy or John's perspective, or a neutral one in between. The actual frame composition and the sequencing of shots varies depending on your choices. In addition, the selection of sounds, as well as their spatial positioning and the mix of sounds varies based on your choices.

Playout from John's Perspective

A cold, barren attic, muted colors. Kathy, a young woman, sits at her typewriter, wind howls as she types, the sound of the typewriter is overpowering. Most of the scene is played out in long shot. John walks up to a dress form and pulls a rag off the form complaining "I'm out all day and I have to come home to *this*?"

Isis

Isis is a multi-purpose, high-level interpretive programming language developed by Stefan Agamanolis. Isis is named after the ancient Egyptian goddess of fertility. The design of Isis is optimized for use as an efficient scripting language for interactive multimedia. Isis is a complete programming language, supporting loops, conditionals, and recursive procedures with a syntax similar to Scheme. Among the primitive data types in Isis is the timeline. Timelines are essential for expressing time and space variant data with linear and cubic interpolation. This feature helped make the code for *Two Viewpoints* efficient and elegant. A comprehensive description of Isis is available in Agamanolis (1996).

Structured Sound

Structured Sound (Inguilizian, 1995) represents sounds as independent audio sources localized in time and 3D space within an acoustic environment. The auditory equivalent to structured video, Structured Sound takes the audio components of an Isis script and interactively renders them with respect to the position of the listener/viewer. The audio components are discrete sounds and effects synchronized with the actions of script objects, while the acoustic modeling and processing performed accounts for the listener location.

Playout from Kathy's Perspective

We hear the faint sound of typing, we see a close up of the typewriter, then a close-up of Kathy. We hear the sound of a gentle wind playing wind chimes. John walks up to the dress form which Kathy sees as Kathryn. John grabs a handkerchief from the dress form, which we see as Kathryn, the character from Kathy's fictional world.

4 The SAR Toolkit

Motivation

Building *Two Viewpoints* in the manner we did was a very laborious process, 'hand crafted' so to speak. Although the development of **Isis** has brought the production environment up a level of abstraction compared to previous structured video productions at the laboratory, making a movie in this manner is much like doing graphic design by writing Postscript code. It's powerful and flexible, but for a moviemaker it's not the optimum level of abstraction for thinking about and previsualizing motion picture sequences.

As we were building *Two Viewpoints* I began designing and implementing some tools, a preliminary attempt to provide an environment in which a moviemaker could develop and experiment with an interactive, multivariant, movie. My primary motivation is to help make the process more flexible and less laborious. An important design criteria is the ability to quickly make changes and see the effect of those changes in the playout under different situations. Moviemakers tell me that one of their primary concerns with new media tools is that the tools should become transparent, 'stay out of the way' and facilitate the flow of the creative process, allowing them to visualize shots and sequences without requiring a long lag time between idea and visualization. One reason the **Avid Media Composer** has been so well received by editors is the ability to see a new sequence without the laborious cutting, splicing and unsplicing that film editing requires, especially when moviemakers want to compare variations.

Scope

Creating a structured video sequence in which the framing is dynamic and the playout is multivariant requires a rigorous methodology in the design, production, and editing phases. A design tool must effectively address these issues. The "Some Assembly Required" toolkit, or **SAR** for short, has been designed to work with already existing structured video elements. In contrast, the **Moviemaker's Workspace** (Higgins,

1994) addressed the issue of previsualizing sequences in the design and preproduction phase.

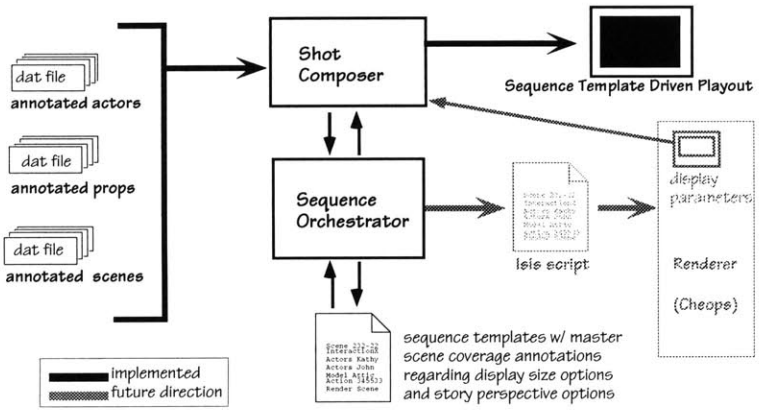
Design Approach

One of the greatest challenges for a designer of storytelling tools is maintaining the underlying narrative structures that tell a *cohesive* story in a manner that remains easy to manipulate for the moviemaker. In this thesis, we are primarily concerned with *viewing context* (the size and aspect ratio of the display) and *viewer's preferences* (which point of view they wish to see) which is inherently limited by the *content available.* Unlike systems that are limited by the clips in their database, structured video provides more options for creating variations on the fly. We avoid the issue of anything being possible, the playout can't be *arbitrary,* because of the narrative structure chosen for *Two Viewpoints.* Since tools developed in a vacuum rarely shed much light on the nature of the production process, I chose to tailor **SAR** to deal with the specific problems of encoding the shots and coverage for *Two Viewpoints.*

The Moviemaker's Activity

The interplay of framing, blocking, sequencing, performance, and point of view is fluid and subtle, and it should remain the moviemaker's choice how and when to 'slant' or 'filter' part of a scene from a camera perspective that reflects a particular point of view, be it through a character, narrator, or omniscient perspective. Moviemakers may use subtle or not so subtle variations in camera position, camera movement, sound, focal length, etc. to 'filter' or 'slant' the scene one way or another.

The **Sequence Orchestrator** essentially supports the construction of a Master Scene Cinema Language template for a scene. This template consists of multiple views or shots defined with **Shot Composer.** The Master Scene Cinema Language, which has become the conventional method for covering a scene, provides for a master shot, over the shoulder coverage, and close ups, all at more or less regular human eye level. This is usually done without unmotivated camera movement. The conventions of Master Scene Cinema Language derive out of the history of representational two dimensional art, and it has gained a kind of authority as an 'objective' viewpoint. As soon as you introduce low or high camera angles, unmotivated camera movement, shift in screen direction, or shots representing inner-subjective experience, you are introducing a new optical point of view, which may have subjective meaning in terms of



The SAR toolkit

Shots are framed up using the *Shot Composer* and cameras are be defined by a series of key frames created with *Shot Composer.*

Using the *Sequence Orchestrator* camera coverage for a scene can be defined on a timeline in terms of which cameras are appropriate for different subject perspectives or display frame sizes.

the story, even if the viewer doesn't recognize it as such. The distinctions of point of view, in the many uses of the phrase, are tremendously complicated. As a cinematographer, these considerations are an integral part of my decision making process as I collaborate in the process of audio-visual storytelling. These decisions have considerable effect on the nature of the viewing experience.

To create a piece that can be viewed from different points of view challenges the moviemaker. What choices can the viewer appropriately make? How can multiple perspectives be expressed cinematically? In order to craft the variations within a story, our tools must support enough flexibility to experiment and try out different variations on framing and editing, and thus craft the variations right in with the story.

The Toolkit

The current toolkit is composed of two programs, **Shot Composer** and **Sequence Orchestrator,** plus a utility, **XmDat,** for taking care of some post-production tasks. The toolkit allows a director, cinematographer, or editor to define synthetic shots and camera movements, as well as narrative constraints in the form of a coverage strategy using Master Scene Cinema Language. Annotations can be added in order to customize shot composition relative to display size. This system points towards better integration between preproduction, production, and postproduction.

Currently these tools allow the maker to previsualize a structured video sequence within the **Open Inventor** environment running on a Silicon Graphics Indigo 2 workstation. An extension of the **Sequencer Orchestrator** to generate an **Isis** script (Agamanolis, 1996) is currently under investigation.

Open Inventor

Open Inventor is an object-oriented toolkit consisting of a library of objects and methods used for developing interactive 3D graphics applications. Open Inventor enables you to write programs that take advantage of graphics hardware features with a minimum of programming effort (Wernecke, 1994).

This would allow a moviemaker to use this toolkit to design shots and sequences with framing and interaction parameters which could be previsualized in the **Open Inventor** environment and then executed on Cheops for the final playout without having to hand-assemble an **Isis** script

An additional tool, **XmDat**, supports a graphics environment for viewing 2D image and structured video files. These tools import structured video elements in the form of DAT files in the strvid format (file formats used at the Media Lab) and thus are compatible with Cheops.

Shot Composer

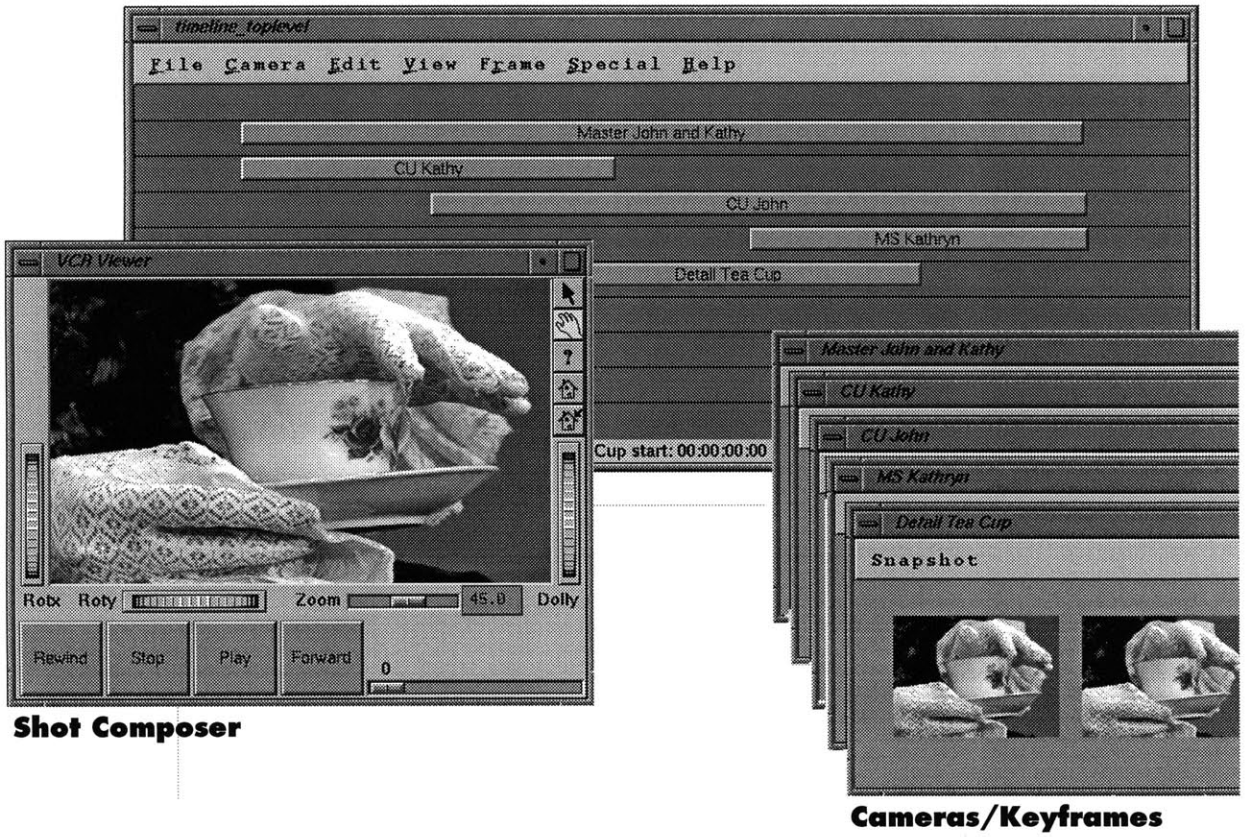
The **Shot Composer** provides the ability to compose shots with either a static or moving camera. It can import structured video elements in the *strvid* file format (an internal file format for structured video in use at the MIT Media Lab).This includes all of the objects and background that we used in *Two Viewpoints,* providing the ability to create cameras and previsualize a scene. When the program is first started it reads the structured video elements specified in a configuration file. The **Shot Composer** import module interprets the 3D data in the strvid file that was created in previous post-production steps with **Scene Analyzer** (Becker, 1995). Each of the elements in the strvid file has also been tagged with a starting timecode number using **XmDat**, therefore **Shot Composer** can play the scene forward and backwards and the proper position of each element is maintained in both space and time. Using a viewer module based on the **Open Inventor** scene viewer, **Shot Composer** allows the maker to move the camera anywhere in the scene. Camera shots can span over any length of time, the tools allows the maker to create a starting 'frame' and an ending 'frame' for a particular camera. If the maker specifies more than two frames for a particular camera, **Shot Composer** will interpolate a path along the key frames specified. This allows the maker to specify moving camera shots in time synchronized with the actions of the video objects in the scene. Each camera is represented as a window with the key frames defining the camera

Sequence Orchestrator

Sequence Orchestrator allows the maker to construct a 'coverage map' of the scene. It is a representation of the temporal and shot coverage relationships in the scene from the perspective of Master Scene Cinema Language. With this map the maker can annotate which cameras are valid during what parts of the scene. The moviemaker can create three edit and camera variations, one for John's POV, one for Kathy's POV and one

that is neutral. In addition, the maker can specify alternative cameras and edits for small, medium and large display systems. If given more than one valid camera at a particular edit point, **Sequence Orchestrator** can either pick a camera randomly or make a weighted choice. For example, on a large display the moviemaker might prefer the wider shot, but on a smaller display they would rather use a more specific shot. Overlapping these two shots in time on the coverage map would then cause **Sequence Orchestrator** to make a choice during playback depending on the current screen size parameter settings.

Sequence Orchestrator



Shot Composer

Cameras/Keyframes

XmDat

XmDat provides the ability to view DAT files and attach annotations to facilitate the tracking of structured video elements through the production process. The tool features a graphical user interface alternative to **xdat**. It supports the annotation of starting timecode, text notes, *ckrects* and area of interest rectangles. The ckrects annotation feature was used to identify the area of interest when we processed DAT files with the chroma-key procedure. **XmDat** allows the user to mark

ckrects and area of interest rectangles on single frames on a frame by frame basis or on a range of frames. Area of interest rectangles may be used in later production step to identify salient objects in the frame. **XmDat** was used for preparing DAT files for the chroma key process by using the ckrect feature which allows the user to mark a rectangle of interest on a frame by frame basis or on a range of frames. In addition, **XmDat** can add various annotations to the descriptor portion of the DAT file including the starting timecode. The starting timecode data is used by the **Shot Composer** and **Sequence Orchestrator** and in order to synchronize multiple views from different cameras upon playout.

Implementation Notes

Structured Video Objects

The **SAR Toolkit** is able to import structured video actors and props which are stored in DAT files in the strvid format. 3D scene descriptions must be in **Open Inventor** format. These file format which are identical to the formats we used for *Two Viewpoints*. The specific post-production steps taken to create these structured video objects and scenes is described in Agamanolis (1996).

Shot Composer

ShotComposer is implemented in C++ using the **Open Inventor** class library from Silicon Graphics, the DAT file I/O library and OSF Motif. The program can read structured video objects stored as DAT files in the *strvid* file format. These are converted internally into the **Open Inventor** format for ease of processing and manipulation in the **Open Inventor** environment.

Sequencer Orchestrator

Sequencer Orchestrator is was implemented in C++ using OSF Motif. It is currently running on a Silicon Graphics Indigo 2 under IRIX 5.2. The **Sequencer Orchestrator** can read and save edit sequences comprised of multiple cameras and coverage annotations. These sequences are stored in a text file with a specific format documented in the source code. In addition, work is currently underway to save a sequence description as a text file in **Isis** script format for playout on Cheops.

XmDat

XmDat was implemented using C and Motif on the DEC Alpha under OSF/3, and it also runs on the Silicon Graphics under IRIX 5.2.

An Integrated Toolkit

The **Sequencer Orchestrator** and **Shot Composer**, originally developed as stand-alone programs, now runs as a integrated unit.

Performance

The **SAR Toolkit** is currently running on a Silicon Graphics Indigo 2 under IRIX 5.2 with the Extreme Graphics option. The performance of this system, as implemented, is quite slow. This is due to the combination of several factors: (1) strvid files are currently read from disk and limited buffering is done, (2) the amount of memory required to read all of the frames of the video objects into memory is prohibitive, and (3) the Indigo 2 is not what I would call a fast machine. A SGI Reality Engine does not improve performance very much. This is partially due to some optimizations which could be done but are not high priority at this time. Even if optimizations would be made, there is still not enough texture map memory available on the lab's Reality Engines to compete with the performance of Cheops if the required structured video files are pre-loaded into the memory buffers on Cheops.

5

Evaluation, Reflections, and Suggestions for Future Work

Two Viewpoints

Two Viewpoints is currently running on the Cheops imaging system. Informal conversations with viewers has been illuminating.

Interaction must be integrated with the narration

One of the things I've observed when showing the system to visitors is that the interactive aspect that engages their interest the most is the ability to change perspectives between Kathy and John. Give the chance to move around in three-dimensions has been described as 'cool,' however, viewers seem more fascinated by 'movement' that has significance in terms of the narrative than just a change in optical perspective.

> The lyf so short, the craft so long to lerne.
>
> — Chaucer

Hardware limitations

The current hardware configuration of Cheops limits the continuous experience to less than two minutes and requires moviemakers to keep the number of media objects as low as possible in order to maintain a decent frame rate. In the end these limitations were not as daunting as we expected. Based on my informal discussions with several viewers, the experience, although short, 'works' as a narrative.

The frame rate, albeit low, is not a major problem

The playout is currently a bit choppy with a relatively low frame rate, however, this step-frame effect actually provides the experience with a touch of surrealism that does not seem to detract from the experience for most viewers.

Is it really a multivariant movie?

Although we call this a multivariant movie, our story yields a more restricted case than the filter stories described in Evans (1994), at least in terms of sequencing. The playout of *Two Viewpoints* is not *arbitrary*, because of the tightly authored narrative structure chosen for the piece. Our goal is to maintain the underlying narrative structures that tell a *cohesive* story. We are primarily concerned with *viewing context* (the size and aspect ratio of the display) and *viewer's preferences* (which point

of view they wish to see) which is inherently limited by the *content available.* Unlike video orchestration systems that are limited by the clips in their database, structured video provides more flexible options for creating variations on the fly, both in terms of montage *and* mise en scène.

How 'smart' is it?

Multivariant movies must contain information on how to *compose* each shot and how to *sequence* the shots in order to build a sequence that maintains narrative coherence. There must be an overall story structure of some kind. In *Two Viewpoints* we implement the ability to shift point of view through the use of **Isis** *timeline structures.* Our tools on Cheops, at least so far, are not in any way a system to understand stories the ways humans do, however, we do encode some of our 'artistic intentions' and provide 'templates' for the construction of sequences based on narrative constraints. More work needs to be done in terms of story structures and knowledge representation.

Alternative Interface

The knob interface of *Two Viewpoints* leaves a little to be desired, as does the mouse and menu interface of the SAR Toolkit. Just as we've scaled the image size to a particular display device, we should explore a variety of interfaces for different playout environments. For example, a multi-modal interface using gesture, speech and eye-tracking could provide a variety of ways for the viewer to navigate through the narrative and express choices (Bers, Elo, Lassiter & Tamés, 1995).

Analog video blues

Despite the fact that a video engineer did everything possible to align the cameras as closely together as possible, the images digitized from the five cameras did not match in terms of colorimetry. Part of this situation was aggravated by the fact that we forced to use different camera models. The new Sony Digital Betacam camcorder, with digital signal processing and 'personality modules,' helps eliminate most aspects of the problem, and I urge future productions to use this camera.

Characters did not blend into the scene in a seamless manner

Some viewers noticed the 'cut-out figure' appearance of Kathy, Kathryn, John, and the props. This was due to the fact that Cheops does not support blending of the image along the edge of the alpha channel mask, known as alpha blending. Adding support for this to Cheops would improve the realism of the images.

Actors and objects did not cast shadows

We did not process shadow information when we did the alpha mask extraction. In addition, Cheops does not provide hardware support for rendering models and applying textures to them in a programmatic fashion (like a Shader in Renderman), therefore, it is difficult to support 3D modelling capabilities like casting shadows onto 3D objects in the scene.

The SAR Toolkit

The SAR **Toolkit**, as it stands today, certainly demonstrates the creative potential and the viability of the approach. We have been able to create sequences that play out differently on small and large screens, as well as representing multiple playouts for different points of view. These multiple playouts are represented as alternative edits using a master scene language template. The tool set is still primitive, and could be extended in various ways.

Export Isis scripts

Currently, sequences only play out on a Silicon Graphics Indigo 2 workstation. A potential enhancement would be to generate an **Isis** script based on the edit decisions represented by the **Sequence Orchestrator** in order to play the final sequences on Cheops.

Extensibility

One possible approach for creating extensions to the toolkit would be to support an **Isis** interpreter (Agamanolis, 1996) in order for users to extend the tool in a programmatic fashion as well as supporting custom programming for specific pieces of content. Another possible approach would be to integrate the **Sequence Orchestrator** and **Shot Composer** with **Isis** itself.

Performance

Playout of sequences on the Silicon Graphics Indigo 2 with the Extreme Graphics option is painfully slow. Solutions to this problem would involve developing an efficient caching scheme for video objects as well as running on a faster platform like a Silicon Graphics Reality Engine.

Support the use of sound

Sound is a critical aspect of moviemaking, however, the SAR Toolkit is still in the silent era. Since the toolkit already supports timecode it would be straightforward to synchronize sound with the structured video elements. Another feature would be adding the ability to do asynchronous sound and picture editing as described in Kung (1995).

Provide more intelligent assistance to the user

The **Sequence Orchestrator** and **Shot Composer** allow for quick previsualizing of dynamic framing and multivariant playout. Currently the tool does not support much beyond play out according to a sequence template. A richer set of annotations and the ability to implement more complex cinematic rules would provide more interesting and subtle variations in playout. One promising approach would be to implement design rules based on relations grammars. This approach been successfully applied to dynamic page layout (Weitzman, 1995). The **Shot Composer** requires that the user decide on key frames in order to implement camera moves. These moves are implemented using linear interpolation between the key frames. A promising next step would be to implement camera moves preserving framing based on cinematic rules. One approach for this is described in Drucker (1994). Another approach for intelligent camera control using vision techniques is described in Pinhanez and Bobick (1995).

More flexibility in the playout algorithm

SAR only implements one rule for dynamic framing, and the support for creating a flexible multiple point of view narrative is limited to what can be represented in terms of a coverage diagram using master scene cinema language. More flexible and generalizable approaches to representing story knowledge that are directly applicable to this toolkit are discussed in Evans (1994), Kung (1995), Murtaugh & Davenport (1995) and Davis (1995).

Structured Video Production Methods

Pre/Pro/Post

Production methodologies using structured video techniques begin to blur the distinction between pre-production and post-production. With *Two Viewpoints* much of the post-production tasks being scripted and actually being carried out at the moment of 'projection.' We begin to see a change in the manner in which media messages are created. Structured video objects with content-specific annotations offer a framework for storytellers to experiment with new story forms. At the same time, this approach makes intense demands on the moviemaker. We are just beginning to scratch the surface with tools and techniques for managing complex relationships within a dynamic story.

Postproduction

The process of shooting, digitizing, chromakey, alpha-channel

mask extraction, 3D data extraction, etc. for *Two Viewpoints* was quite tedious and required a large number of manual steps. **XmDat** took a stab at facilitating some of the steps, but much remains to be done. If we could automate these intermediate postproduction steps such that they could be accomplished in a matter of hours instead of weeks, the design and production of movies using structured video techniques could compete favorably on an economic basis with traditional video production.

Reflective Moments

What I love about cinema, and one reason I'm a cinematographer, is the ability of cinema to weave together a complex, multidimensional story and deliver it in a dream-like two hour stream of consciousness experience that gets assimilated into my consciousness. Great literary and cinematic experiences expand my awareness of many things in delightful ways. Not a single interactive title I've experienced so far has left a major mark in my consciousness, yet many films, like *Apocalypse Now, Annie Hall, Manhattan, Blade Runner, Citizen Kane, Speaking Parts, Meshes of the Afternoon,* etc. have become part of my consciousness. I'm able to integrate their 'texts' into my own reflection on experience and they help explain and make sense of my experiences. These films are more than just entertainment, they say something about life, values, what it means to be human, etc. Good art helps clarify and refine human experience, challenging us to see things from a different point of view. I'm not saying that the intense aesthetic and emotional effect of cinema is not possible with interactive media, however, I've not seen it yet. As a moviemaker I look forward to the day I can take back my statement about interactive media. I understand that evolving a new media form takes a long time and a great deal of work. Cinema started as a curiosity over a hundred years ago. It's time to venture out into the world and see what I can do about it.

A1 *TwoViewpoints* Original Scenerios

John

John comes home from work tired. He finds that the house is in disarray and goes up to the attic. He finds Kathy who is seated at a desk writing. He berates her angrily for not cleaning, etc. She ignores him completely and continues staring at her typewriter. He leaves, slamming the door. He returns to the kitchen and stares away. His anger leaves him and he begins to cry. He breaks his glass in frustration.

Kathy

Kathy sits at a desk with a typewriter. She stares at her page. She types for a moment, and stares off again. She looks at the wallpaper in the room. A face seems to appear, faintly. She looks back ad her page. John bursts into the room and yells at he for not cleaning the house. He screams at her but she does not respond. The face appears again in the paper and she stares at it. John leaves. Kathy gets up and strolls to the paper. She feels it. She looks back towards the page. There is a bright light. There is a high jingling sound. A lacy gloved hand sets a tea cup on a saucer. The hands feels the material of a satily dress. The hand passes slowly over a cobblestone wall. There is a loud crash. Kathy looks up suddenly from her desk. The room is very quiet and Kathy stares off again.

Interaction

The viewer can see the ongoing sequence from either Kathy or John's perspective. The actual dialog and character actions remain the same, however, a different subjective position, or focalization, is created by the choice of cuts, the framing, and the manner in which the images are composited together. Elements of the scene will consist of the room, the wallpaper pattern, Kathy, Kathryn (the 19th century character in Kathy's story world), John, a dress form, lacy gloves, a teacup and saucer, a cobblestone wall, and a glass.

A2 Two Viewpoints Original Script

Kathy's Perspective

Written by Tony Romain

INT. ATTIC - DAY

KATHY sits at a desk in a bright attic. White light streams in from a large window. Kathy stares out the window and then down to a typewriter on the desk in front of her. Dust moves lazily in the stream of sunshine over her desk. It is quiet.

MAN
(O.S.)
Kathy! Kathy!

Kathy turns very slowly towards the voice then back to her desk. There is a faint ringing sound off in the distance.

The door to the attic opens loudly, JOHN walks in. His face is obscured somewhat by shadows. He walks quickly over to Kathy.

JOHN
What the hell is going on here? Can't you hear me yelling down there?

Kathy turns away from John and smiles. She looks over towards a corner of the room. There is a dresser's dummy perched in a corner. She looks back towards the window then back to the corner. There is a now a woman that looks like Kathy there. She is wearing a long flowing 19th century gown. She holds her laced gloved hands in her lap eloquently. She has a subtle smile on her face. She pulls a handkerchief from the fold of her dress and dabs her eyes. She then puts it neatly back in her lap. John walks over to her and pulls the handkerchief from her and shoves it in Kathy's face.

The gloved hand of the woman sets a teacup on a

saucer. She feels the fabric of her gown, it makes a soft purring sound as she moves her hand.

There is a crashing sound.

Kathy looks up suddenly from the typewriter. She is alone in the room. She looks towards the door and then back to the window.

John's Perspective

Written by Tony Romain

INT. KITCHEN - DAY

JOHN walks tiredly into a messy kitchen. There are dishes everywhere and food remains strewn about. He takes off his coat and looks around the kitchen incredulously. He sighs deeply.

JOHN
Kathy! Kathy!

He looks up to the ceiling knowingly. He turns quickly and walks up the stairs.

INT. ATTIC - CONTINUOUS

John opens the door to the attic sharply. It is filthy. Dust streams from the ceiling over old books and crumpled papers. A dresser's dummy lies disheveled in a corner a dirty rag is draped over it. In the center of the room sits KATHY. She is hunched over a typewriter and barely notices John's entrance.

JOHN
What the hell is going on here? Can't you hear me yelling down there?

Kathy mumbles something. She keeps her gaze locked on her page.

JOHN
What have you been doing all day? This place is filthy!

John walks in front of the desk Kathy is sitting at. She looks up briefly at him then returns to her page.

KATHY
I don't know.

JOHN
Jesus, I'm out all day, why would I want to come home to this?

John paces around the room. He picks up the rag off of the dummy and looks at it. It is fetid and moss covered. Kathy's gaze is unmoved from her page.

JOHN
(exasperated)
Look around you!
What's wrong with you?!

John throws the rag down and leaves the room, slamming the door behind him.

INT. KITCHEN - CONTINUOUS

John walks back into the kitchen angrily. He slams his fist into a table. His anger leaves him and he looks sadly up towards the ceiling. He moves his hand and knocks over a mug on the table. He looks down surprised and bends down to pick up the pieces.

A3 *Two Viewpoints* Shot List

0 Calibration for Master

1 Master shot, Kathy and John

2 Master shot, Kathy only

3 Master shot, John only

4 Dress Form

5 Calibration for Tracking Shot

6 Tracking shot, Kathy, Dress Form, and John

7 Tracking shot, Kathy only

8 Tracking shot, Kathryn only

9 Tracking shot, John only

10 Tracking shot, Dress Form only

11 Calibration for Medium Shots

12 Medium shot, John and Kathryn

13 Medium shot, Handkerchief/Rag

14 Medium shot, John and Kathryn/Dress Form

15 John

16 Kathy

17 ECU Kathy

18 ECU Teacup and Kathryn

19 ECU Kathryn

20 Dress and Kathryn

21 ECU Handkerchief

22 John in Kitchen

23 John breaks coffee mug in Kitchen

Shots in italics made it into the final cut of *Two Viewpoints,* the scene with John in the Kitchen was cut completely.

A4 Blue Screen Compositing

Lighting for *Two Viewpoints* was pretty standard, although the challenge was to match the lighting of the original room. Far cycs for the walls, scoops for overall coverage, zips for fill, and a 5K fresnel as a key to match the window light of the set. We painted the cyc Chroma Key Blue.

The blue screen technique is a proven method for compositing objects (usually actors) into backgrounds. The actors can be shot in a studio and later composite into a three-dimensional model. The chromakey process analyses the RGB values of each pixel in the original image and creates a corresponding alpha channel in which the object mask is represented as black pixels. Levels of gray can be used for smooth alpha blending between model and object and compositing shadows into scenes.

It is especially important when going to digital video to make sure that the lighting on the blue screen is a consistent hue and saturation of blue. I've used both chromakey blue and ultimatte blue, and my experience, when doing the digital compositing in software, the chromakey blue provided cleaner results with less problems due to variations in the hue. With *Two Viewpoints* we discovered that some scenes had inconsistent hues of blue, which would have 'passed' with Ultimatte, were exaggerated in our rgb digitized video dat files. Some of the more serious problems were fixed by either creating a custom blue-patch reference that included these off hue blues for use as a reference with our in-house *ultra-chromakey* program or by editing the alpha channel by hand. Not fun.

In retrospect the best contemporary solution for doing blue screen work is the Ultimatte System, which allows you to see your compositing against your background during the shoot. This allows you to optimize your lighting and camera adjustments with less guesswork. For the next Cheops shoot, I suggest having a digitizing workstation on the set and use the Ultimatte System, or at least our in-house *ultra-chromakey,* on the set so you know right there and then if you have a good alpha channel or not.

The talent should be as far away from the blue screen as possible, preferably 12 feet or more. You must be careful whenever

using colored gels on foreground lights, as spill from these could change the hue of the blue screen. Watch out for the color of the talent's clothing and hair. Light colors that reflect the blue are a problem. Use a back light to eliminate any blue bounce is one solution and provide better definition against the background. Sometimes green screen is the answer. Objects that are to be composite can be lit from any angle, and shadows cast onto the blue screen can be transferred onto the alpha channel. Make no assumptions. On *Two Viewpoints* Kathy was wearing tan pants. Not a good idea, and thus the strange artifacts on her legs in some views of *Two Viewpoints.*

Some of the blue is extracted from the foreground objects in the chromakey process, so if this causes a problem, there is a technique called color difference mask for getting this color back into the foreground objects.

References

Agamanolis, Stefan. "High-level scripting environments for interactive multimedia systems." M.S. Thesis, Massachusetts Institute of Technology, 1996.

Almendros, Nestor. *A Man with a Camera.* New York: Farrar, Straus, Girou, 1984.

Arijon, Daniel. *The Grammar of the Film Language.* Hollywood: Silman-James Press, 1976.

Becker, Shawn. "Vision-assisted Modeling for Model-Based Video Coding," Ph. D. Thesis Proposal, Massachusetts Institute of Technology, 1995.

Becker, Shawn, and V. Michael Bove, Jr. "Semiautomatic 3-D Model Extraction from Uncalibrated 2-D Camera Views." *SPIE Symposium on Electronic Imaging: Science & Technology,* February, 1995.

Bers, Josh, Sara Elo, Sheri Lassiter & David Tamés. "CyberBELT: Multi-Modal Interaction with a Multi-Threaded Documentary," *CHI '95 Companion,* Denver, Colorado, May, 1995.

Balázs, Béla. *Theory of the Film,* translation of *Filmkultura,* trans. Edith Bone. New York, Dover Publications, 1970.

Bender, Walter, V. Michael Bove, Jr., Andrew Lippman, Lin Liu, & John Watlington. "HDTV in the 1990s: Open Architecture and Computational Video," *HDTV World Review* 1.3, 1990, 11-15.

Benson, K. Blair & Jerry Whitaker. *Television and Audio Handbook.* New York: McGraw Hill, 1990.

Boorman, John & Walter Donohue, eds. *Projections: A Forum for Film Makers,* Issue No. 1. London: Faber and Faber, 1992.

Boorman, John & Walter Donohue, eds. *Projections: A Forum for Film Makers,* Issue No. 2. London: Faber and Faber, 1993.

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

Bordwell, David. *Making Meaning: Inference and Rhetoric in the Interpretation of Cinema.* Cambridge: Harvard University Press, 1989.

Bove, V. Michael Jr., Brett Granger and John Watlington. "Real-Time Decoding and Display of Structured Video." In *IEEE ICMMCS '94,* Boston, Massachusetts, 1994, 456-462.

Bove, V. Michael. Jr. and Watlington, John. "Cheops: A Reconfigurable Data-Flow System for Video Processing," *IEEE Transactions on Circuits and Systems for Video Technology,* 5(2), 1995.

Bove, V. Michael, Jr. "Synthetic Movies Derive from Multi-Dimen-

sional Image Sensors." Ph.D. Dissertation, Massachusetts Institute of Technology, 1989.

Branigan, Edward. *Point of View in the Cinema: A Theory of Narration and Subjectivity in Classical Film.* Berlin: Mouton, 1984.

Branigan, Edward. *Narrative Comprehension and Film.* New York: Routledge, 1992.

Brecht, Bertolt. *Brecht on Theatre: The Development of an Aesthetic.* ed. and trans. by John Willett, London: Eyre Methuen, 1964.

Bresson, Robert. *Notes on the Cinematographer.* trans. Jonathan Griffin, London: Quartet Books, 1975.

Brinton, Joseph P.. "Subjective Camera or Subjective Audience?" *Hollywood Quarterly* 2, 359-65, 1947.

Browne, Nick. "Introduction." *Film Reader* 4, 1979, 105-7.

Browne, Nick. "The Spectator-in-the-Text: The rhetoric of *Stagecoach.*" *Film Quarterly* 34(2), 1980.

Browne, Nick. *The Rhetoric of Filmic Narration.* Ann Arbor: University of Michigan Press, 1982.

Bruckman, Amy. "The Electronic Scrapbook." M.S. Thesis, Massachusetts Institute of Technology, 1991.

Bruner, Jerome. "Two Modes of Thought," in *Actual Minds, Possible Worlds.* Cambridge, Massachusetts: Harvard University Press, 1986.

Davenport, Glorianna, Thomas Aguierre Smith and Natalio Pincever. "Cinematic Primitives for Multimedia." *IEEE Computer Graphics and Applications,* July, 1991, 67-74.

Davenport, Glorianna, Ryan Evans and Mark Halliday. "Orchestrating Digital Micromovies." *Leonardo* 26(4), 1993, 283-288.

Davenport, Glorianna. "Seeking Dynamic Adaptive Story Environments," *IEEE Multimedia* 1(3), 1994, 9-13.

Davenport, Glorianna. "Still Seeking: Signposts of Things to Come," *IEEE Multimedia* 2(3), 1995, 9-13.

Davenport, Glorianna and Michael Murtaugh. "ConText: Towards the Evolving Documentary." in *Multimedia '95,* San Francisco, California, 1995, 381-389.

Davis, Marc. "Media Streams: Representing Video for Retrieval and Repurposing." Ph.D. Thesis, Massachusetts Institute of Technology, 1995.

Drucker, Steven Mark. "Intelligent Camera Control for Graphical Environments." Ph.D. Thesis, Massachusetts Institute of Technology, 1994.

Eisner, Will. *Comics and Sequential Art.* Tamarac, Florida: Poorhouse Press, 1986.

Elliott, Edward L. "Watch • Grab • Arrange • See: Thinking with Motion Images via Streams and Collages." M.S. Thesis, Massachusetts Institute of Technology, 1993.

Elsaesser, Thomas, Ed. *Early Cinema: Space - Frame - Narrative.*

London: British Film Institute, 1990.

Evans, Ryan. "Log Boy meets Filter Girl." M.S. Thesis, Massachusetts Institute of Technology, 1993.

Fielding, Raymond, ed. *A Technological History of Motion Pictures and Television.* Berkeley: University of California Press, 1967.

Fielding, Raymond, *The Technique of Special Effects Cinematography,* 4th ed. Boston: Focal Press, 1985.

Fiske, John. *Television Culture.* New York: Methuen, 1987.

Foley, James D., Andries van Dam, Steven K. Feiner and John F. Hughes. *Computer Graphics: Principles and Practice.* Reading, Massachusetts: Addison-Wesley, 1990.

Friedberg, Anne. *Window Shopping: Cinema and the Postmodern.* Berkeley: University of California Press, 1993.

Freuder, Eugene C. and Richard J. Wallace (1992). "Partial constraint satisfaction," *Artificial Intelligence* 58, 21-70.

Galyean, Tinsley. "Narrative Guidance of Interactivity." Ph.D. Thesis, Massachusetts Institute of Technology, 1995.

Gaudreault, Andre (1987). "Narration and Monstration in the Cinema." *Journal of Film and Video* 39, 29-36.

Gilman, Charlotte Perkins. *The Yellow Wallpaper.* Thomas L. Erskine and Connie L Richards, eds. New Brunswick, New Jersey: Rutgers University Press, 1993.

Granger, Brett D. "Real-Time Structured Video Decoding and Display." M.S. Thesis, Massachusetts Institute of Technology, 1994.

Higgins, Scott. "The Moviemaker's Workspace: Towards a 3D Environment for Previsualization." M.S. Thesis, Massachusetts Institute of Technology, 1994.

Haase, Kennneth, B. "Framer: A Persistent Portable Representation Library." *ECIA-94 Proceedings,* March 1994.

Halliday, Mark. "Digital Cinema - An Environment for Multithreaded Stories." M.S. Thesis, Massachusetts Institute of Technology, 1993.

Houbart, Gilberte. "Viewpoints on Demand: Tailoring the Presentation of Opinion in Video." M.S. Thesis, Massachusetts Institute of Technology, 1994.

Inguilizian, Araz. "Synchronized Structured Sound: Real-Time 3-D Audio Rendering." M.S. Thesis, Massachusetts Institute of Technology, 1995.

Johnson, Michael Boyle. "WAVESworld: A Testbed for Constructing 3D Semi-Autonomous Animated Characters." Ph.D. Thesis, Massachusetts Institute of Technology, 1995.

Karp, Peter and Steven Feiner. "Issues in Automated Generation of Animated Presentations." In *Graphics Interface,* Halifax, Nova Scotia, 1990, 39 - 48.

Katz, Steven. *Film Directing Shot by Shot.* Studio City, California:

Michael Wiese Productions, 1991.

Katz, Steven. *Film Directing Cinematic Motion.* Studio City, California: Michael Wiese Productions, 1992.

Lasky, Alan. "Slipstream: A Data-Rich Production Environment." M.S. Thesis, Massachusetts Institute of Technology, 1990.

Laurel, Brenda. *Computers as Theatre.* Reading, Massachusetts: Addison-Wesley Publishing Company, 1991.

Lehnert, Wendy G. "Plot Units and Narrative Summarization." *Cognitive Science* 4, 1981, 293-331.

Lehnert, Wendy G., Malcolm E. Cook, and David D. McDonald . "Conveying Implicit Content in Narrative Summaries." Department of Computer and Information Science, University of Massachusetts, Amherst, 1984.

Lehnert, Wendy G. "Narrative Complexity Based on Summarization Algorithms." Computational Models of Natural Language Processing. B.G. and G. Guida Bara, ed. B.V. (North-Holland): Elsevier Science Publishers, 1984.

Mandler, Jean. *Stories, Scripts, and Scenes: Aspects of Schema Theory.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1984.

Mamet, David. *On Directing Film.* New York: Viking Penguin, 1991.

Mathias, Harry and Richard Patterson. *Electronic Cinematography.* Belmont, California: Wordsworth Publishing Company, 1985.

McCloud, Scott. *Understanding Comics.* Northhampton, Massachusetts: Kitchen Sink Press, 1993.

Meehan, James. "The Metanovel: Writing Stories by Computer." Computer Science Research Report 74, Yale University, 1976.

Minsky, Marvin. "A Framework for Knowlege Representation." AI Lab Memo 306, Massachusetts Institute of Technology, 1974.

Minsky, Marvin. *The Society of Mind.* New York: Simon and Schuster, 1986.

Monaco, James. *How to Read a Film,* revised ed. New York: Oxford University Press, 1981.

Moreno, Julio L. "Subjective Camera: And the Problem of Film in the First Person." *Quarterly of Film, Radio and Television* 7, 1952, 341-58.

Morgenroth, Lee. "Movies, Talkies, Thinkies: An Experimental form of Interactive Cinema." M.S. Thesis, Massachusetts Institute of Technology, 1992.

Mullet, Kevin & Darrell Sano. *Designing Visual Interfaces: Communication Oriented Techniques.* Englewood Cliffs, New Jersey: SunSoft Press, Prentice Hall, 1995.

Negroponte, Nicholas. *Being Digital.* New York: Alfred A. Knopf, 1995.

Oldman, Gabriella. *First Cut: Coversations with Film Editors.* Berke-

ley: University of Califonia Press, 1992.

Pinhanez, Claudio S. and Aaron F. Bobick. "Intelligent Studios: Using Computer Vision to Control TV Cameras." *IJCAI '95 Workshop on Entertainment and AI/Alife,* 1995.

Prince, Gerald. *A Dictionary of Narratology.* Lincoln: University of Nebraska Press, 1987.

Richards, Ron. *A Director's Method for Film and Television.* Boston, Massachusetts: Focal Press, 1992.

Rosenblum, Ralph and Robert Karen. *When the Shooting Stops, The Cutting Begins.* New York: Viking Press, 1979.

Rubin, Ben. "Constraint-Based Cinematic Editing." M.S. Thesis, Massachusetts Institute of Technology, 1995.

Sayles, John. *Thinking in Pictures.* Boston: Houghton Mifflin Company, 1987.

Schank, Roger and Robert Abelson. *Scripts, Plans, Goals and Understanding.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Schank, Roger and Robert Abelson. "Knowledge and Memory: The Real Story." in *Knowledge and Memory: The Real Story. Advances in Social Cognition, Volume VIII* (series), Robert S Wyler, Jr., ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1995.

Schubin, Mark. "The History of the Perfect Aspect Ratio," in *Proceedings of the 137th SMPTE Technical Conference,* New Orleans, September 6-9, 1995, 459-482.

Stam, Robert, Robert Burgoyne & Sandy Flitterman-Lewis. *New Vocabularies in Film Semiotics.* London: Routledge, 1992.

Toulet, Emmanuelle. *Cinematographe, invention du siecle.* Paris: Gallimard/Réunion des Musées Nationaux, 1988.

Truffaut, Francois. Hitchcock, Rev. ed. New York: Simon and Schuster, 1984.

Tsang, Edward. *Foundations of Constraint Satisfaction.* London: Academic Press, 1993.

Weitzman, Lois. "The Architecture of Information: Interpretation and Presentation of Information in Dynamic Environments." Ph.D. Thesis, Massachusetts Institute of Technology, 1995.

Wernecke, Josie. *The Inventor Mentor.* Reading, Massachusetts: Addison-Wesley Publishing Company, 1994.

Wilson, George M. *Narration in Light: Studies in Cinematic Point of View.* Baltimore, Maryland: John Hopkins University Press, 1986.

Winston, Patrick Henry. *Artificial Intelligence.* Reading, Massachusetts: Addison-Wesley, 1992.

Wood, Sharon. *Planning and Decision Making in Dynamic Domains.* New York: Ellis Horwood, 1993.

Youngblood, Gene. "Cinema and the Code." *Leonardo,* Computer

Art in Context Supplimental Issue, 1989, 27-30.

Zettl, Herbert. *Sight, Sound, Motion: Applied Media Aesthetics.* 2nd ed. Belmont, California: Wadsworth Publishing Company, 1990.

"At the end of a life spent in the pursuit of knowledge Faust has to confess: 'I now do see that we can nothing know.' That is the answer to a sum, it is the outcome of a long experience. But as Kierkegaard observed, it is quite a different thing when a freshman comes to the university and uses the same sentiment to justify his indolence. As the answer to a sum it is perfectly true, but as the initial data it is a piece of self-deception. For acquired knowledge cannot be divorced from the existence in which it is acquired."

— Dietrich Bonhoeffer,
Cost of Discipleship