

If You Could See What I Mean...  
Descriptions of Video in an Anthropologist's Video Notebook

by  
Thomas G. Aguierre Smith

B. A. Social Science  
University of California  
Berkeley, California

1989

Submitted to the Media Arts and Sciences Section, School of Architecture and Planning,  
in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE  
AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
September 1992

COPYRIGHT MIT, 1992  
All Rights Reserved

Signature of Author

Media Arts and Sciences Section  
August 1992

Certified by

Glorianna Davenport  
Assistant Professor of Media Technology,  
Asahi Broadcasting Corporation Career Development Professor,  
Director of the Interactive Cinema Group

Accepted by

Stephen A. Benton  
Chairperson  
Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

NOV 23 1992

LIBRARIES

If You Could See What I Mean...  
Descriptions of Video in an Anthropologist's Video Notebook  
by  
Thomas G. Aguierre Smith

Submitted to the Media Arts and Sciences Section, School of Architecture and Planning on  
August 7, 1992  
in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

**ABSTRACT**

The Anthropologist's Video notebook is a video database application that allows researchers to present movies in a format that reflects the contextual complexity of ethnographic data. The Anthropologist's Video Notebook is grounded in both the practice of ethnographic research and motion picture production. The lexical descriptions of video content are represented using the Stratification system. Stratification is a context-based layered annotation method which treats descriptions of video as objects. Stratification supports the complementary and at times contradictory descriptions which result when different researchers use video source material which is available on a random access video workstation. The development of the Anthropologist's video notebook is based on a real field work experience in the state of Chiapas Mexico. The integration of ethnographic research methods and video production heralds a new research methodology called video ethnography. Video ethnography is the study of how meanings are attributed to video over time. The Stratification system allows for the analysis of the significance of the content of video in terms of the context of where it was recorded and also the context where it appears in an edited sequence.

Thesis Supervisor:

Glorianna Davenport  
Assistant Professor of Media Technology,  
Asahi Broadcasting Corporation Career Development Professor,  
Director of the Interactive Cinema Group

If You Could See What I Mean...  
Descriptions of Video in an Anthropologist's Video Notebook

by  
Thomas G. Aguierre Smith

The following people have served as readers for this thesis.

Reader:

.....

Kenneth W. Haase, Jr.  
Assistant Professor of Media Arts and Sciences

Reader:

.....  
Brent Berlin /  
Professor of Anthropology and Integrative Biology  
University of California at Berkeley

## ACKNOWLEDGMENTS

I would like to thank Janet Cahn, Eddie Elliot, Stephan Fitch, Natalio Pincever, Laura Teodosio, Stanley Sclaroff and David Yategari for their friendship and intellectual playfulness.

In the Interactive Cinema Group, I thank Erhhung Yuan and Joshua Holden the two undergraduate researchers who have diligently worked on the implementation details of the Stratification modules. I have also enjoyed working with Hiroshi Ikeda and Hiroaki Komatsu, two research affiliates in the Interactive Cinema Group from the Asahi Broadcasting Corporation. I wish much success to Lee Hayes Morgenroth, Mark Halliday and Ryan Evans.

Out side of the lab, Andy Podell has given me intellectual and moral support. Andy provided much needed reality checks and for this I am grateful. I thank Ken Haase. And in Mexico, I thank Brent and Eloise Ann Berlin, and the researchers of the PROCOMITH project for sharing their enthusiasm for integrating audio visual media into ethnographic research.

And finally I would like to thank Glorianna Davenport for believing in me and creating a stimulating research environment that enabled me to explore the ideas that are presented in this thesis.



## TABLE OF CONTENTS.

Abstract .....	2
Acknowledgments .....	4
Table of Contents. ....	5
1. Introduction .....	7
2. The Medium, the Method and Antecedents for the Video Notebook .....	10
2.1 Making Observations, Interpretations and Descriptions .....	10
2.1.1 Inscription with Video .....	13
2.1.2 Text and Ethnographic Movies .....	15
2.1.3 Can You See What I Mean? .....	16
2.1.4 Digging through Context.....	16
2.1.5 Multimedia Ethnography with Digital Media .....	18
2.2 Antecedents for the Anthropologist's Video Notebook .....	19
2.2.1 The Validity of Video Data.....	22
2.2.2 Issues for Describing Content in an Anthropologist's Video Notebook.....	24
3. Making Memories of Video Content in a Field Setting .....	25
3.1 Finding a Method that Works -- Developing the Anthropologist's Video Notebook .....	26
3.1.1 Strategy One: Segmentation .....	27
3.1.2 Strategy Two: Content Markers .....	29
3.1.2.1 Keywords and Context .....	31
3.1.3 Strategy Three: Stratification .....	32
3.2 Beyond Logging: Video as Resource .....	35
4. Representation and Description in Motion Picture Production .....	37
4.1 An Image is a Description .....	39
4.2 The Descriptive Dimensions of a Video Stream .....	40
5. Audio Visual Descriptions by Design .....	41
5.1 Toward a Theory of Video as Design.....	42
5.2 Experimenting in a Design Space .....	45
5.3 Locating Content in a Design Space .....	47
5.4 A Design Environment for Shooting Video .....	49
5.4.1 Two Types of Context .....	50
5.4 Design Environment for Editing Motion Pictures.....	53
5.5.1 Stratification and Virtual Video Editing .....	57
6. Making Knowledgeable Lexical Descriptions .....	60
6.1 Moving Image Databases and the Ghost of St. Augustine. ....	61
6.2 Language Games -- The Fibers and "Strata" of Understanding .....	62
6.3 Implications for the Automatic Logging of a Video Stream .....	65

7. Making The Anthropologist's Video Notebook.....	67
7.1 The Field Notebook: System Configuration .....	67
7.2 Logging Video with CueTrack .....	69
7.2.1 CueTrack Stack Modifications .....	70
7.3 The Keyword Stack .....	71
7.3.1 Using the Keyword Stack .....	72
7.4 The Sequence Stack.....	73
7.4.1 Using the Sequence Stack .....	74
7.5 Making Translations .....	74
7.6 Getting the Big Picture.....	76
7.6.1 Of Time Codes, Video Editing and Switching Computer Platforms .....	76
7.7 The Stratification System - A Design Space for VIDEO .....	79
7.7.1 System Configuration .....	79
7.8 Data Representations.....	80
7.8.1 Strata Data Format .....	80
7.8.2 The UNIX File system .....	82
7.8.3 Keyword Classes.....	82
7.9 The Stratification Modules .....	85
7.9.1 GCTL - Galatea Controller .....	86
7.9.2 VIXen: The Logger.....	87
7.9.2.1 Free Text Description .....	89
7.9.2.2 Keyword Classes .....	89
7.9.2.3 Annotation Techniques .....	90
7.9.3 Stratagraph.....	92
7.9.3.1 Modifying the Display.....	95
7.9.3.1.1 Real Time Display. ....	95
7.9.3.1.2 Scale of Difference .....	96
7.9.3.1.3 Segmentation and Collapsing Keywords .....	97
7.9.4 Making Sequence Annotations: Infocon .....	99
7.9.4.1 Picture Icons - Picons .....	101
7.10 Naming a Sequence as Data Entry .....	102
8. Conclusion .....	104
References .....	106

## 1. INTRODUCTION

The research presented in this thesis reflects video observation and other forms of field notation which were conducted while engaged in anthropological field work in the state of Chiapas, Mexico. The development of a computer representation for video called *stratification* reflects the demands of an anthropologist who must organize and present a content base of video taped observations. Stratification is a context based annotation method that treats lexical descriptions of the video stream as objects. The content of the video stream can be derived by looking at how these descriptive layers or strata overlap and relate to each other. Stratification was used to create a video database system called the Anthropologist's Video Notebook.

The Anthropologist's Video Notebook is modeled after the anthropologist's field notebook. Field notebooks are collections of descriptions, impressions, and data that represent an anthropologist's attempt to understand a situation as it unfolds. The notebook is used as a recording medium. It allows the researcher to take observations out of context and away from the time and place where they were recorded. The impressions and memories that surround these "captured" observations are used to interpret and organize these field notes with the final aim of producing more detailed descriptions that will communicate what happened to someone who was not there.

Ethnographic data is inherently multimedia. The principle design issue for an Anthropologist's Video Notebook is how to technologically support the integration of ethnographic research and movie production. Ethnographers have used audio visual media to record observations but the publication format for anthropology had been text. Observations are made in real time and then written down in a notebook. These notes are then arranged and transformed into monographs, lectures, books, etc. Although text is static (observations become fixed in a text), these observations become dynamic *over time* as new information becomes incorporated during the later stages of writing. The observational movie maker uses video tape rather than a notebook to fix observations. As with the anthropologist's notebook, video is a dynamic medium which can be reinterpreted in a myriad of ways. Not only does the content of video change *in time* (frames are recorded at a rate per second), it changes *over time* as chunks of video are arranged into sequences.

The development of an anthropologist's video notebook is grounded in the understanding of how anthropologists and movie makers work with a medium in order to communicate what they have observed to other people. There is an interplay between the medium and the method. When the media of text and video become integrated in a random access video database system, a new research methodology of video ethnography emerges. The key issue is to understand how the computer can function as a

representation system that mimics the way in which anthropologists and movie makers interact with media to produce new content. For both the anthropologist and the movie maker, data gathering occurs within the context of time and place. Observations of people (what they do and say) have a temporal and spatial extent. Events occur in a moment and must be remembered, interpreted and retold. Both anthropologists and movie makers respectively use the media of text and video<sup>1</sup> to record and represent dynamic information.

The design of the anthropologist's video notebook requires that we treat the practice of video ethnography as an information processing task where content changes *in time* and how that same content changes as it is used and re-purposed *over time*. Ricki Goldman Segall in *Learning Constellations*, explains: "The questions most relevant to video ethnography are: how do researchers reuse what they see on video to make serious generalizations; and, how does the content -- the meaning of what happened -- remain intact. In other words, how can we rely on our video data? As in other approaches, the issue is one of rigor in one's method of extraction." And later she warns, "... the dilemma facing builders of new technologies is how to give the user freedom of movement within the source material while maintaining the integrity of the original event" (Goldman Segall, 1990: p 17, 50).

An integrated representation for the content of video is needed. Such a representation must maintain the descriptive integrity of the source material while supporting the re-juxtaposition of the "raw video" in to new edited contexts. A *design environment* for video ethnography needs to be created that supports the interplay between the rigorous analysis of video data and the maintenance of the original descriptions. The former is a prerequisite to the generation of new theories and a deeper understanding about a culture while the latter allows for descriptive coherency over time and throughout the production process - from shooting to editing.

Chapter Two is a discussion of how the development of the Anthropologist's Video Notebook is grounded in the process of creating an ethnographic text and the process of making a movie. These two processes come together on today's computers that integrate both video and text into one system. In Chapter Three, the problems of working with video in a field situation are explored. Stratification is presented as a way to represent the content of a video stream. In Chapter Four, the way that a particular representation scheme affects the types of descriptions that are possible is discussed in regard to the Stratification method. The way that Stratification can be used to create a information environment or design environment for video ethnography is covered. Chapter Five emphasizes video as a design process while Chapter Six focuses on the problem of ambiguous lexical descriptions of images. The nuts and

---

<sup>1</sup>All references to video could be extended to other type of temporally based media.

bolts of implementing the Anthropologists Video Notebook are presented in Chapter Seven. And in Chapter Eight, I conclude.

## 2. *THE MEDIUM, THE METHOD AND ANTECEDENTS FOR THE VIDEO NOTEBOOK*

The development of the Anthropologist's Video Notebook involves thinking about the ways that both ethnographic data and observational style movies are produced. Specifically, how are the modes of recording observations constrained by the traditional media of ethnography (text) and observational movies (video)? The medium effects how actions and utterances are recorded and how this recorded content is later interpreted and transformed into a new type of content - a final work. The respective mediums of text and video enable ethnographers and observational movie makers to experiment by arranging and juxtaposing concepts and images that are based in observed reality.

We need to explore on a deeper level the creative process behind ethnography and motion picture production to see how they can be effectively coupled using a computer system. The computer can provide a new medium for the integration of ethnographic research practice and video production. But before this integration can take place it is necessary to examine how a written text and a video tape function not only as descriptive media but also as interpretive media. A written ethnography or an ethnographic film are interpretations of things that were observed.

### 2.1 *Making Observations, Interpretations and Descriptions*

Clifford Geertz in *The Interpretation of Cultures* (1973) characterizes ethnographic description as an interpretive practice. He explains,

So, there are three characteristics of ethnographic description: it is interpretive; what it is interpretive of is the flow of social discourse; and the interpreting involved consists in trying to rescue the "said" of such discourse from its perishing occasions and fix it in perusable terms (Geertz, 1973 : 20).

In simpler terms, the ethnographer makes interpretations which are based on observed actions and utterances as they unfold at particular time and place. The ethnographer fixes his interpretations so that they can be "perused," analyzed and studied in a different place and time and by people who were not present to witness the actual event. By fixing an interpretation the ethnographer rescues an observation from the moment and place where it was made.

How does the ethnographer accomplish the task of fixing an interpretation? Geertz rhetorically asks and answers an analogous question: "What does the ethnographer do? -- he writes" (1973: 19). The anthropologist gains understanding by first creating and then working with textual descriptions. Geertz further elaborates:

The ethnographer 'inscribes' social discourse; *he writes it down*. In doing so he turns it from a passing event which exists only in its own moment of occurrence, into an account, which exists in its inscriptions and can be reconsulted (Geertz, 1973 :19 original emphasis).

Writing is the principle mode of representation for anthropology and the compilation of notes is the primary activity when doing ethnographic fieldwork. Yet, the anthropologist Simon Ottenberg points out, notes are not necessarily the only traces of an interpretation:

There is another set of notes, however, that anthropologists consider to be incorporeal property. These are the notes in my mind, the memories of my field research. I call them my head notes. As I collected my written notes, there were many more impressions, scenes, experiences than I wrote down or could possibly have recorded (Ottenberg, 1990: 144).

The impressions that have been committed to memory are an important and yet fragile component of the interpretive practice of anthropology. The anthropologist must rely on his memory throughout the research process. The initial notes that are written at the scene of an observation serve as a memory aid or mnemonic for later stages of creating written descriptions. Ottenberg gives a good account of the how memory and writing evolve during research:

My fieldnotes themselves are based upon "scratch notes" taken in long hand with a pen on small pads of paper and then typed up in my "free time" -- often in the late evening when I was quite fatigued. The handwritten notes are brief sentences, phrases, words, sometimes quotes -- a short hand that I enlarged upon in typing them up, adding what I remembered. ... So my hand written notes are my original written text, and my typed notes are a reinterpretation of the first interpretation of what was in my head when I produced my handwritten notes (Ottenberg, 1990: 148).

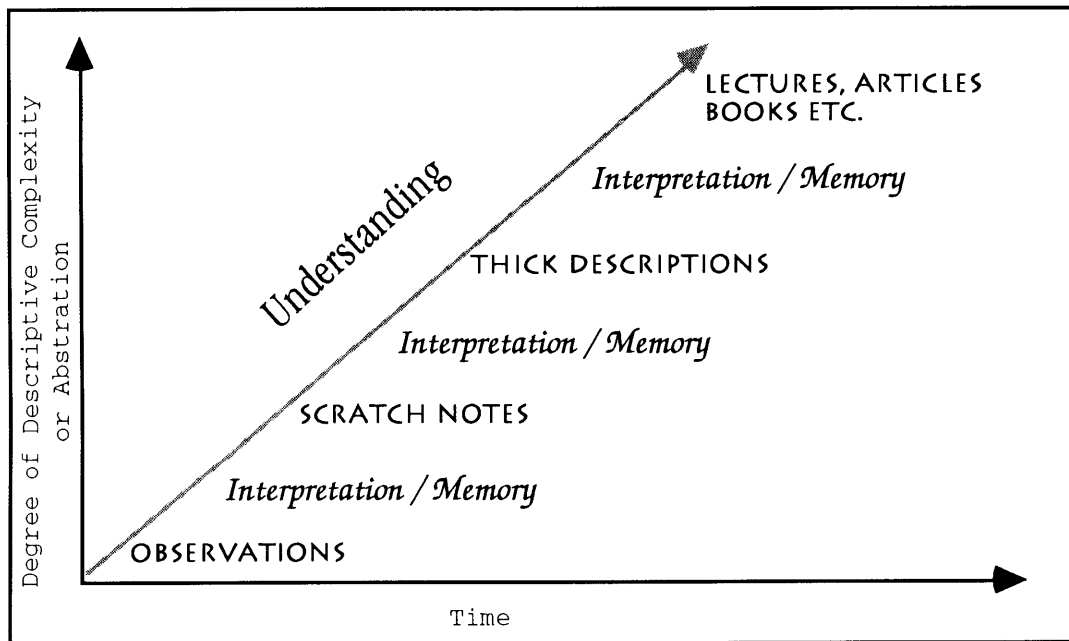
The initial written notes are referred to as "Scratch notes." Scratch notes represent the moment when the anthropologist's perceptions and interpretations are first committed to paper. At a later time, and upon reflection of the anthropologist's memory, the scratch notes are transformed into more formal typed notes. For Geertz these "enlarged" upon notes are called *thick descriptions*. In creating a thick description, the anthropologists must tease out and incorporate the background elements of observed events -- the insider information -- into a coherent description for the reader. Geertz, explains that this background is essential for the communication of significance to an outsider:

In finished anthropological writings ...this fact --that what we call our data are really our own constructions of other people's constructions of what they and their compatriots are up to -- is obscured because most of what we need to comprehend a particular event, ritual, custom, idea, or whatever is in-

sinuated as background information before the thing itself is directly examined (Geertz, 1973: 9).

A thick description maintains the contextual coherency of an observed action or utterance. It brings the background into the foreground and cements the researcher's fleeting memory of what happened into the ethnographic text. Thick descriptions are a coherent representation of observed cultural reality and serve as a database for later writing and interpretation that is aimed at the production a finished account. Written texts of thick descriptions can be contrasted, compared and juxtaposed in different ways. They serve as the raw material that the ethnographer uses to question, speculate and theorize about the significance of things observed. A thick description is the embodiment of the ethnographer's memory of context. Throughout the research process the anthropologist's memory can easily be inscribed into a text. The composition of a text from remembered observations is the key feature of field work (figure 1).

Figure 1: Complexity of Ethnographic Description Over Time. Ethnographic content evolves as research progresses. During each stage of writing the text serves as a representation system. Over time memories and new interpretations are integrated into more abstract and complex textual descriptions.



Although the text remains fixed at any given moment, the process of creating textual descriptions of observed actions is dynamic because it changes over time. The process of writing gives the ethnographer much latitude for representing and manipulating interpretations and memories of observations. The creation of thick descriptions evokes the memory of the context of where and when significant actions and utterances occurred. At all phases during ethnographic research the anthropologist's memory can



be incorporated or “fixed” into a text. Furthermore, these texts can be used in a variety of new contexts to produce written monographs, articles, and even lecture notes. The textual content of ethnographic research is in a constant state of transformation as the research progresses.

### *2.1.1 Inscription with Video*

Video is a medium that is used by observational movie makers to describe observations and experiences. Where the ethnographer produces scratch notes, the video maker simultaneously fixes observations of events and utterances onto the medium of video tape while observing them. He uses a camera to record significant actions and utterances onto the video tape. Yet, to see how video can be truly integrated into the interpretive practice of ethnography, we have to take a close look at how the video maker works with the medium and appreciate the role that memory plays in this process.

Lev Kuleshov illuminates some of the challenges of working in a medium of motion pictures:

In the cinema the understanding of the material and the understanding of the organization of the material are particularly complex, because the material of the cinema itself demands particular organization, demands particularly extensive and specifically cinematic treatment. The cinema is much more complicated than other forms of art, because the method or organization of its material and the material itself are especially “interdependent” (Levaco, 1974 : 188).

The content of video evolves out of two interdependent processes, namely the process of shooting and editing. The material (a series of contiguous images exposed on film) and the organization of the material (montage or editing) are interdependent because the makers compose the content of the frames while shooting and then take these artifacts into the editing room where they are composed or ordered into sequences. Kuleshov’s characterization of working with the medium raises some key issues.

Although the process of recording observations is analogous to ethnographic data gathering, the way that the material is shaped during production differs. The ethnographer can enlarge upon his scratch notes by incorporating memories and new interpretations while writing. In this regard, it is the memory of the event which provides the source material that is interpreted and evaluated as research progresses. For the observational video maker, the inscribed bits of video *become* the memory of the event. The observational video maker can only work with things that have been inscribed/ remembered on the medium. In other words, the source material for an observational movie maker are the actions and utterances which occurred while the camera was on. The stakes are higher for the video maker. The way that he has interpreted the action as it unfolds has ramifications for later stages of movie produc-

tion. The movie maker Richard Leacock instructs us to use the camera to actively interpret the scene as it unfolds:

Use the camera as an extension of your own eye. You look you search. Yes, you pan and tilt. You learn to look for light and position yourself to take advantage of its beauty. You learn to sense the movement of the camera in space. You think of the image not merely as a way of showing something but also as a way of withholding information, of creating tension in the viewer. Of not revealing too much. Of seeing things with different perspectives by using different focal-length lenses. Collecting images that are not obviously connected, to create a space for your own making. And also it goes, with the caveat that nothing counts until you see it on the screen. You look and look until you know it is right. (Leacock, 1990 : 3).

Although a moving image is an artifact of the process of making a recording at a certain place and time, it can be re-purposed during editing. The reuse by cutting and reordering /editing (also called montage) of contiguously recorded frames is one of the most intriguing features of cinema. In the 1920s, Kuleshov created an editing experiment that demonstrated what is now known as the Kuleshov effect. The experiment consisted of alternating “the same shot of Muzhukhin (the actor) with various other shots (a plate of soup, a girl, a child’s coffin), and these shots acquired a different meaning.” He explains, “The discovery stunned me — so convinced was I of the enormous power of montage” (Levaco, 1974 : 200). The Kuleshov effect best illustrates how disparate chunks of video can be ordered to produce a new meaning that transcends the original intent of each individual piece. Kuleshov went so far as to write that the reordering of image artifacts was the most important type of cinematic content. He claimed that what was “conventionally thought of as the narrative or dramatic “content” of a film was irrelevant to the structuring of this “material” (Levaco, 1974 : 7). Kuleshov’s irreverence to narrative content reveals the seduction of how the medium of film allowed the makers to depict cinematically constructed events and actions out of the careful reordering of recorded images.

When we think of the role that memory plays in ethnographic research and in video production it seems that they are at odds with each other. However, the observational film maker is not unlike the ethnographer. Where the ethnographer inscribes his interpretations in a text by writing in a notebook, the observational video maker inscribes his interpretations on a video tape by recording sounds and images with a camera. Geertz could easily conceive of the process of shooting a video as the cinematic inscription of an event as it unfolds. Inscription occurs at the moment that something significant is perceived and shot by the video maker. The recorded images become the video maker’s memory - if the camera was not recording when something occurred then there are no moving images that can be used in the final movie. The process of making an observational film is analogous to writing an ethnography with only the words that were used in the notebook while the event was being observed. The production

of the final video is dependent on images that have been recorded, if it was not recorded then it can not be used in an edit. On the other hand, the anthropologist's memory plays a critical role by filling in the blanks. When creating thick descriptions, the anthropologist uses his scratch notes as mnemonic devices which aid in the production of rich descriptions. In this way, the anthropologist's memory becomes integrated into the final work. For the video maker, memory is auxiliary to the construction of the final edited work. In the editing room the video maker reflects upon his memory of experiences and tries to project how an outsider will perceive a certain configuration of the raw footage in an edited context. The memory of the video maker guides the production of the final movie but it is external to medium, the recorded images are the core. The video maker's memory is critical to the construction of a meaningful sequence. Yet, this memory is external to images recorded on a video tape.

Where the production of an ethnographic text represents the integration of the contextual memory into a textual medium, movie production can be characterized as the struggle to maintain the contextual memory of what was shot. A video tape often exists without the video maker's memory. As Kuleshov has noted, moving images have an amnesia that allows them to be re-used in a variety of contexts. This poses some challenges for the ethnographic researcher who decides to use video.

### *2.1.2 Text and Ethnographic Movies*

The way that anthropologists typically use video in their research illustrates the difficulty they have had in working with the medium. The difficulty and the expense of making a visual record coupled with the lack of tools to annotate and incorporate one's contextual memory into visual media are some of the reasons why text has remained the traditional analytical and publication format for many anthropologists. The lack of tools does well to explain why many of the films that are created in the context of anthropological research are used like other cultural artifacts - an accompanying text is usually written to explain the events in the film. Text is the only presentation format that gives researchers the flexibility they require to associatively browse their data and transform it into various publications. Although the computerization of databases has streamlined this process for analysis, the presentation of results has still largely been limited to the "hard copy" medium.

Many anthropologists typically trans-code video into other more manageable formats. For example, stills are created from long movie sequences. In most cases these key frames are further reduced to textual description which finally end up in a written report (Collier, 1986). A projective interviewing technique which used film was developed by Worth and Adair (Worth & Adair, 1972). In this pioneering study, Navajos made films of their own choosing. The goal was to see if the way that the Navajo structured their films could be a reflection of the Navajo world view. Krebs (1975) screened films with special time indices for informants to elicit social norms evident in a Thai dance drama called Khon. Bellman

and Jules Rosette (1971) analyzed how Kapelle camera men approached the creative process of shooting video. Their primary interest was to study how they intentionally used the medium to depict events and actions of religious rituals. Yet, in all of these cases, motion pictures were the artifacts that were analyzed and written up in the final report. Although the use of the medium was the object of study, the motion pictures were relegated to the appendix as an illustration of points raised in the written report.

### *2.1.3 Can You See What I Mean?*

Although the video tape is an inscription of an action or utterance there is no way to link the textual descriptions which reflect the maker's memory of context or changes in interpretation to the video tape itself. Some of the frustrations of working with video in an anthropological research setting can be seen in one of my past research projects.

In 1989, I studied how pedestrians crossed a busy intersection as a research project in Professor John Ogbu's field methodology course in the Anthropology Department of the University of California at Berkeley. In conducting the research I found that traditional ethnographic techniques of note taking would not be sufficient in capturing the fast paced action and events of crossing the street. Due to the difficulty in observing whole acts I decided to use video. I thought that the use of the medium of video would help me decipher the complex interactions and grammar of street crossing behavior. After repeated viewing, I made a log or description of the tapes in written form. The more I watched the tape the better I was able to identify different styles of street crossing behavior. I came to see the video with more understanding. The subtleties of the pedestrians actions became clear.

When I edited together a short tape of my findings and showed the tape to the class, they were disoriented - they could not see what I meant. The sequences of street crossing behavior were a reflection of my intent to record pedestrians in a situated environment and a reflection of my analysis of what types of activities were significant. Yet the class could not see the interactions in the same way that I had seen and interpreted them. It was difficult for a new viewer to perceive the subtleties that I had gained through repeated viewing. To get around this problem, I produced a written explanation of my findings and a final research report called Negotiated Crossings (Aguierre Smith, 1990).

### *2.1.4 Digging through Context*

I had thought that the video tape would be a valuable analytical tool for the study of complex social scenes instead it turned out to be a cryptic form of representation that needed to be supplemented with a text. Something is wrong with the current modes of recording observations with video. To get a deeper understanding of the problem of using video it will be useful to approach the motion pictures recorded on the video tapes like an archeologist who traces the significance of excavated artifacts. Yet

unlike the archaeologist, the ethnographer who uses video as a recording medium is in the role of the maker of the artifact. And as the maker of this video artifact he is in the awkward position of doing archeology in reverse. Video is shot within a certain context. Throughout the production process the contextual coherency video artifacts must be maintained. The ethnographic video maker has two responsibilities. First he is responsible for the creation of cultural artifacts - the actions and utterances that I record with my camera. And second and perhaps more important, he is responsible for the “excavation” of meanings that are contained in these artifact and communicating these meanings to other people. We can better understand the problem from an archaeological framework. Let’s examine how the archeologist works with contextual information:

*Archaeologists George Stillson and Linda Towel, the supervisor of the archaeology branch of the National Park Service’s North Atlantic Region, discuss the value of context while working on a dig in Eastham, Massachusetts (McLaughlin, 1990 ).*

— You could take all the artifacts from this or any other site, throw them in a cigar box and put them on my desk, and 95 percent of their meaning would be lost.

— Of course artifacts are important and we are finding cutting and chipping tools, hide scrapers, bits of pottery and so forth. But, it’s the context in which they’re found that makes all the difference. We look for characteristics of the soil, changes in color and texture, pieces of animal bone, charcoal, seeds, even traces of pollen that don’t emerge until we’re in the laboratory. And more ephemeral data, ‘features’ we call them, like hearths, post-hole molds, tool-working areas. Context is everything in archaeology.

To the archaeologist, contextual clues are a critical aid in the discovery of what an artifact represents. The frustration of the archaeologist upon finding a bunch of artifacts in a cigar box is akin to the frustration my classmates felt upon viewing my raw footage. They both lacked the contextual knowledge with which to make sense of the artifact. In other words, they could not have a memory of where it was found.

Given a finite amount of video material (a two hour video tape) how do I go about finding a particular shot or sequence? The method that I use and the success of the search depends on the relationship that I have had with the material. When I view a video tape that I recently recorded, I compare what I remembered about the flow of events (as they actually occurred) with the order of shots as the video tape is played back. I use this personal knowledge or “memory” of what was going on while I was recording to guide me to the particular shot that I’m interested in. My memory of the context within which I shot the video aids me in finding the chunk that I want.

As viewer who has never seen the video before, you can not recall the flow of events as they actually occurred. Your memory can't provide sufficient contextual information. If you find yourself in this predicament, it is advantageous to go back to the time when the video tape was recorded. In this best of all possible worlds scenario, you would have known the following information: where the camera was; what and/or who the camera was pointed at; how the image was framed in the view finder; when the record button was pressed. Most important, the motivation and intent behind making the recording would be evident because you were there.

The process of trying to figure out what an image represents on a video tape has strong parallels with the process that an archaeologist goes through when he uncovers an artifact. For the archaeologist, the artifact is surrounded by environmental features or ephemeral data. The environmental context is a critical factor in the understanding of an artifact. Context is the only information that archaeologists have. We can approach the problem of describing the content of a video tape in the same way that the archaeologist looks at the artifact within the context of where it was found. For the video maker, the context within which the moving images were recorded is analogous to the ephemeral data which surrounds the archaeologist's artifact. The contextual factors that were in effect during recording (the "where", "who", "what", "when", "why" and "how") are the essence of video maker's memory. Without this contextual memory, the images on the video tape are like those curious artifacts in a cigar box.

The question to ask is: Why do we have to treat images like mysterious objects from the past when we have just recorded them? The film is not as mysterious as the archaeologist's artifact. The movie maker knows a lot about the images as he is recording them. Knowledge about the content of the video tape is at its maximum while being recorded and drops to its minimum while being viewed.

### *2.1.5 Multimedia Ethnography with Digital Media*

On the computer, the medium of text which is elemental to the practice of ethnography and the medium of video which is elemental to the practice of observational cinema *collapse* into one media type -- digital. On a computerized multimedia system, *textual descriptions* of actions and utterances will be incorporated with the *audio visual descriptions* of the same significant actions and utterances. Although the text and video are both digital, their content is generated by distinct descriptive practices. The production of content on a random access computer system, such as the Anthropologist's Video Notebook, requires a new descriptive methodology that is the hybrid of the process for working with textual descriptions and video taped recordings of actions and utterances. The textual descriptions will serve as an index into the randomly accessible video. The digital information stream will consist of dynamically changing visual and lexical descriptions.

The development of digital computers which can support both textual and audio visual media promises to transform anthropological fieldwork. Textual descriptions should serve as a collective memory of the content of video. Using the computer we can keep track of lexical descriptions of content that are linked to the video stream. Furthermore, on the basis of these initial logs we can create new sequences - new descriptions of events and actions which reflect the researcher's growing knowledge of the topic. These descriptions are managed by a database. Different researchers with divergent research agendas will describe and use the video footage in regard to their own needs. The method is exploratory - the ethnographer/video maker reflects on the current state of content and then incorporates memories and new insights into the creation of new content. These requirements entail that the Anthropologist's Video Notebook support how maker's interact with the medium as they use their memory of context to not only interpret the significance of content but also to transform content into something new.

## *2.2 Antecedents for the Anthropologist's Video Notebook*

Glorianna Davenport's *A City in Transition: New Orleans, 1983-86* (Davenport, 1987) and Ricki Goldman Segall's *Learning Constellations* (1990) are examples of multimedia systems where textual descriptions are fused to video. In many ways, Davenport's and Goldman Segall's systems are the foundation for the Anthropologist's Video Notebook.

*A City in Transition*, was an experiment in the form of a cinematic case study to discover what kind of movies could benefit from having supplementary documentation available "on-line" for the viewer. It incorporated three hours of movie sequences on laser disc that could be played back via computer control, a still frame library consisting of images of characters, places, and maps, and text annotations that were integrated into a graphical computer display. The system was a watershed in many ways because it not only incorporated non-motion picture elements in a single graphical display but it also allowed for the viewer to investigate how this information was related on his own.

Davenport was particularly interested in how the inherent non-linearity of the medium of computer controlled laser discs in combination with other media elements such as text, and still images impinged upon the movie maker's methodology. She explains:

As an observer, film maker or viewer, I am curious to learn why and how a given individual does what s/he does, and how the cumulative effect of multiple interactions changes the nature of our environment. However, given the traditional limits of linear viewing, the two observational roles -- that of movie maker and movie viewer -- are fundamentally different. . . .

I will call our original recorded observations or first level of interpretation. A second level of interpretation is introduced when we edit our observation into a story. Traditionally the physical linearity of this second level of

interpretation has defined an inviolate temporal experience which generates a third level of interpretation of any viewer. As viewers, we watch, experience and think about the movie but are unable to take an active role in shaping our discovery. (Davenport, 1987 : 4)

The traditional linear form of movies doesn't really allow the viewer to go "back stage" to study a specific element in more detail. Although much background research goes into the production of a documentary it is difficult to incorporate this information into a the linear format of traditional cinema. In this way, *A City in Transition* was an environment that not only engaged the viewer as an observer of the documentary movie but also moved him to become a participant in the exploration of the material that was available on the system.

A key design issue was how to structure the information stream for an interdisciplinary audience. A database was created so that the different types of researchers could find the footage that they would be interested in. The viewer could retrieve sequences of shots, text and images by querying this database. The role of the observer was that of a detective who discovered relationships between information that had already been entered into the database. The more the viewer knew about the content the deeper he could explore the material. He would know what queries to the database would produce an interesting movie. The structuring of content information required that the viewer should know what to look for before hand. Although the viewer could create reports that reflected their own exploration of the material these *were not* re-integrated into the content database. The discoveries made by one individual could not be shared by other users of the system.

These notions are addressed in Goldman Segall's *Learning Constellations*. Her goal was to create a research environment "wherein thick descriptions are created not by the multimedia ethnographic researcher but by the user as well" (Goldman Segall, 1990 : 50). In this system, the overhead of creating the database of textual descriptions that is a gateway into the video sequences was shared by the viewer.

Goldman Segall shot over 70 hours of video while conducting ethnographic research about children's epistemologies at Project Headlight, the Logo constructionist culture that had developed at the Hennigan Elementary School in Boston, Massachusetts. She focused on three children and created a multimedia case study about each.

Goldman Segall's analysis of her video footage mimicked the ethnographic analysis of observational data. Using Apple Computer's hypermedia application Hypercard, she designed an application called "Star Notes" that allowed her to log or describe the video using textual descriptors. Her initial logs are analogous to the ethnographic field worker's scratch notes. Each card of the Hypercard stack



“Star Notes” represented a chunk of video tape. The information included on each card consisted of the video tape name, beginning frame and ending frame, title, free text description and keywords. Keywords served as a way to “enlarge” or “thicken” the visual description that had been inscribed on video tape.

While categorizing each chunk of video, no predetermined set of keywords existed. They emerged from the data. We would watch the video and decide the keywords according to the content (Goldman Segall, 1990: 88).

The keywords were not predetermined but were created as the video footage was analyzed. As text, the keywords provided a way to incorporate new interpretations and memories of the context of where the footage was shot. Later, these keywords with the aid of search and sort routines could be easily compared and contrasted. During this stage of research, the keywords provided the descriptive hooks for the development of content themes. Goldman Segall relates that “themes and thick descriptions guided the exploration and established the constraints within which we negotiated the videodisc real estate” (Goldman Segall, 1990: 92). The themes led to the selection of a representative sample of the video material to be placed on a set of video discs that, in turn, served as the raw material for the multimedia case studies. The selection of video was critical. The thick lexical descriptions informed the selection and assembly of “thick” video sequences that would be the source material for the multimedia case studies. For Goldman Segall, thick descriptions took on a new meaning in the context of video editing:

In editing the video data, both the material and the remembered products come together to create different video slices of the initial experience. Many directions emerge -- each with its own truth to the total experience; each adding a more textured or thick description of what was going on at the time.  
...

What gets placed beside a segment of children’s fingers typing at a keyboard could lead to a piece about the exploration of typing styles by placing a shot of other children’s fingers in the following segment; or it could lead to a piece about body language of children using the computer-- by placing shots of different parts of the body videotaped from different angles, directions and distances in the following video segments (Goldman Segall, 1990: 33).

The raw footage reflects the ethnographers’ understanding of the situation at the time of recording. Portions of the raw material were organized into thick visual descriptions so that an outside viewer could experience a coherent representation of the original observed event or utterance. The material that ended up on the final laser discs was the result of Goldman Segall’s interpretations.

Another Hypercard application called “Learning Constellations” was developed to aid other researchers in the analysis of thick visual descriptions. Learning Constellations was a tool that enabled

the viewer to explore the video in much the same way as an anthropologist explores her data when returning home from the field. It provided a gateway into Goldman Segall's interpretation for the uninitiated viewer. In turn, the user could create personalized annotations of video that were actually linked to the sequences on video disc. In this way, the system could store the user's interpretation of Goldman Segall's interpretation of events and utterances.

What researchers can do in this hypermedia environment is to show thick slices of what they observed. In other words, they can select and share material to let the user come close to their interpretation of the intention of the person who experienced the event (Goldman Segall, 1990: 34).

In this way, thick descriptions could be built up by the viewer although the initial chunking of the information was done by the researcher. Goldman Segall's approach is evocative of Kuleshov's editing experiments - her concern is that the ethnographers show thick slices of what they observed. In editing, one can create context by juxtaposing one image next to the other. Thickness evolves within the temporal axis of the video tape (chunks of video can be arranged into thick descriptions).

### 2.2.1 *The Validity of Video Data*

Goldman Segall points out there are risks for this type of methodology. She explains that by providing viewers with thick slices of video "protects the data from being taken out of context. An alternative chunking is available to users. One can build a grouping of mini-chunks, defining the beginning and end point of each chunk, by using the Video notes builder -- a videodisc editing tool. Researchers who are concerned about having their research re-structured by users need to consider how this affects the interpretation of their data by others" (Goldman Segall, 1990: 99). Thick descriptions are a way to protect against misinterpretation:

Geertz's notion of *thick descriptions* can be used for investigating the meaning and intention of what was being said or done. This search for understanding the meaning of an event often leads to problems of interference with what is going on as well as the personal bias in reporting it. The more the researcher participates, the greater the possibility of interfering with what is happening. ... I raise these concerns in order to address the problem of misrepresentation and misinterpretation of non-linear video. In non-linear multimedia ethnography the pieces of video chunks can more easily be taken out of context (Goldman Segall, 1990: 37).

Goldman Segall is concerned about the implications of allowing the viewer to have access to the raw footage. What happens to the ethnographer's authority if the viewer has access to the first level of interpretation? Such issues have been hotly debated in anthropology (Clifford & Marcus, 1986 and Sanjek, 1990). Like it or not, every anthropologist confronts these issues during field work. Geertz ex-

plains that such conflicts are elemental to the practice of interpretive anthropology and talks about how thick descriptions are also built up over time:

The fact is that to commit oneself to a semiotic concept of culture and an interpretive approach to the study of it is to commit oneself to a view of ethnographic assertions as, to borrow from W. B. Gallie's by now famous phrase, "essentially contestable." Anthropology, or at least interpretive anthropology, is a science whose progress is marked less by a perfection of consensus than by a refinement of debate. What gets better is the precision with which we vex each other (Geertz, 1973: 29).

A program that protects the data from de-contextualization also inhibits re-contextualization and the formation of new and unanticipated understanding. When shooting a video, the images are taken out of context -- the medium is inherently decontextualizing but this freedom of association is what allows makers to create new content out of previously recorded images. The dynamism of the medium should not be squashed. In fact, how the significance of the content of a chunk of video mutates during sequence assembly is characteristic of the development of a thick description. The essentially contestable characteristic of lexical descriptions needs to be extended for visual descriptions that have been recorded on a video tape. Geertz does not mean that thick descriptions preserve once and for all the ethnographer's initial interpretation of an event or that a thick description is a way to objectively preserve observations. Instead he conceives thick descriptions as central to the interpretive practice of anthropology where ethnographic content evolves. Geertz further explains:

Although one starts any effort at thick descriptions, beyond the obvious and superficial, from a state of general bewilderment as to what the devil is going on -- trying to find one's feet -- one does not start (or ought not) intellectually empty-handed. Theoretical ideas are not created wholly anew in each study; as I have said, they are adopted from other, related studies, and, refined in the process, applied to new interpretive problems. If they cease being useful with respect to such problems, they tend to stop being used and are more or less abandoned. If they continue being useful, throwing up new understandings, they are further elaborated and go on being used (Geertz, 1973: 27).

Although the computer becomes the domain where textual descriptions and digital video become integrated, the process of creating an ethnographic text and the process of editing a video tape are distinct. The Anthropologist's Video Notebook integrates a naming environment where memory-based contextual knowledge can be built up for a video, and a video production environment where moving images can be initially logged and used for many different purposes. The task is to figure out how to best establish the link between descriptions of an event which are lexical and descriptions of the same

event which are audio visual. Furthermore, the naming and production space must allow for the creation of “essentially contestable” descriptions of content.

### *2.2.2 Issues for Describing Content in an Anthropologist's Video Notebook*

The leading research issue is not only how to represent dynamically changing content information in a video database system but also how to describe video in such a way that these descriptions can serve to generate new content.

The observations and interpretations of events and utterances which are played out in an environmental context become inscribed in a medium. For the traditional anthropologist the medium is a notebook and type written manuscripts. For the observational movie maker, the medium is video tape or film. In both cases, the process of inscription is inherently decontextualizing. Yet, when combined with the maker's memory, a new type of context can be created with the medium which aims to communicate what occurred to someone who was not present at the scene.

In random access video database systems the way that recorded/inscribed events are selected and presented affects the type of descriptions and conclusions that other researchers can make while working with the material. How does the process of decontextualization and re-contextualization that is possible with a medium affects the type of content that is possible? To tackle this question we must appreciate how a memory --which is inscribed in a textual or visual description of content-- is pulled by the video ethnographer's intention to make an observation and how this contextual/background information of “being there” becomes integrated into a final work. On a random access digital computer the video stream becomes a *resource*. The key design issue is how to represent not only dynamically changing content in time but how to represent content as it changes over time as it becomes reused in new sequences.

### 3. MAKING MEMORIES OF VIDEO CONTENT IN A FIELD SETTING

During the summer of 1991, I spent eight weeks working with Programa de Colaboración sobre Medicina Indígena y Herbolaria (PROCOMITH) in Chiapas, México (Berlin et al., 1988). Project PROCOMITH is a collaborative scientific research program between the State of Chiapas, Mexico and the University of California. Investigators include anthropologists, linguists, botanists, ethnobotanists, illustrators, and pharmacologists from Mexico and the United States. In short, the goals of PROCOMITH are:

- The determination of which plant species the indigenous Maya people of Highland Chiapas attribute with medicinal value, development of a detailed ethnobotanical description of the cultural knowledge of the medicinal uses of these species and finally the evaluation of the pharmacological benefit of these plants.
- The identification and description in biomedical terms of the traditional illnesses which can be treated in the home with plants and the attainment of a comparative description between the Biomedical and Maya medical systems.

The goal for the summer was to video tape various aspects of PROCOMITH's research and organize this material in a video database called the Anthropologist's Video Notebook. By the end of the summer I had shot 27 hours of video. These hours of video tape included footage of how researchers carried out their investigations, visits to rural villages, shots of traditional healers making herbal prescriptions as well as scenes of Mayan Indians living and working in an urban setting.

The video that I recorded of PROCOMITH research is an reflection, an artifact, of my personal goals, desires and intentions to capture something on tape together with the fact that I was at a certain place at a certain time. The shots that I recorded are my cinematic interpretations of events as they unfolded. Unlike my previous experience with my peers at UC Berkeley, when I showed what I had shot to the other researchers who were working on the project, they were intrigued by this form of data gathering. For example, a botanist wanted to know what type of plant was used in a particular shot and a linguist wanted to look for examples of a particular verb tense. All this interest in my work was both exciting and frustrating. I discovered that the video tapes which are "artifacts" of my experience were being contested by their research agendas. I needed to describe the video so that when I returned to the Media Lab I could find shots to make sequences. It also seemed desirable to create a database that all the different researchers could use. But this was impossible: I lacked the knowledge and expertise to rigorously describe the video in all the different disciplines. Moreover, they lacked the context - my *memories* of the actions and events that I recorded. The tension between the creator of an ethnographic videotape

and the viewer of these tapes is embodied in the Anthropologist's video notebook in particular and most video database systems in general.

Although the digital computer provides an environment where textual memories about the context of a chunk of video can be directly attached to the video tape and utilized in a production environment, the development of a computational approach for linking up these memory-based descriptions to the video signal required much experimentation.

### *3.1 Finding a Method that Works -- Developing the Anthropologist's Video Notebook*

The annotation method for the Anthropologist's Video Notebook was developed in an actual field work situation. I was shooting video every day and generating a large amount of video material. The different methods that I explored were developed on the spot and tested. If they did not work, they were quickly rejected. I realized that the more time I spent logging the video tapes the less time I could spend out shooting. The descriptive method had to be fast - I needed to get my descriptions attached to the video tape while my memory was still fresh. Furthermore, the descriptions also needed to be complete so that I could find what I wanted in the editing room. Another concern was the wear and tear on the video tapes, the logs had to be created with only one or two passes.

I found that in creating a video database system where it is possible to access individual frames randomly, each frame needs to be described independently. The researcher needs to represent the atoms (frames) from which chunks of video can be made. When working with a random access video system the application designer must consider the following set of implementation or design issues:

- If the video is segmented from the start, how then can the descriptions support the needs of other users who have to access the video represented in a database for their own purposes?
- What are the objects that we need to describe in a video database? Are they frames, shots, sequences, movie titles or some combination of all four? How can we describe these objects uniquely so that they can be retrieved independently?
- Moreover, how can the representation of content be "elastic" to support various granularities? By this I mean, if there is a chunk of video that consists of 3000 frames (ten minutes) and all I need is 10 seconds how can the computer help me retrieve the best 10 seconds. I don't want to have to look at all ten minutes -- this would be like looking at an unmarked video tape!
- On the other hand does this mean that we have to represent the content in every frame, and in a digital system, every pixel? If we strictly annotate moving image content does the expressiveness of the medium become strangled?

- And finally, how does the way that an image is used in an edited context affect its meaning? Does not the edit itself constitute a new form of content which has to be remembered?

These issues were considered while developing the video notebook in the field. The context based annotation method of video content called “Stratification” emerged as I implemented and tested different logging strategies. In the next three sections the different logging strategies that I used while working in Mexico are presented in chronology. This mode of presentation underscores how the limitations of a given strategy were addressed in the later attempts. It also shows how the stratification method evolved out of the rigors of making and managing video fieldnotes.

### *3.1.1 Strategy One: Segmentation*

The first strategy for building a video data base system in the field involved breaking down the entire video tape into discrete chunks called segments. Segmentation is the technique that was used by both Davenport and Goldman Segall in their systems. Segmentation requires the maker to break down the video into chunks and then to annotate these chunks with textual descriptions. But in defining what the chunks are, the maker must confront with the “granularity problem.” The granularity problem is best illustrated with an example. The following examples are drawn from the field work experience in Chiapas, Mexico.

I tried to use segmentation to rigorously describe a typical street scene in the town of San Cristobol. In this scene a street cleaner who is a young boy walks up the street and stops to pick up some trash in the gutter. Right when he bends down to scoop up the trash, a girl walks down the sidewalk. She is carrying a bright yellow bunch of bananas as she passes right in front of the camera. The boy picking up the trash and the girl walking down the street were happening simultaneously. In the video database I wanted to be able to retrieve the whole shot of the street cleaner walking up the street, scooping up the trash and dumping it into the trash barrel. I also wanted to retrieve the frame with the girl carrying the bananas. The shot of the “girl with bananas” is contained in the shot of the “boy street cleaner.” These two shots overlap each other.

I needed to annotate what was unique about each frame in order to retrieve it independently. The amount of description and the work required to distinguish one frame from the others was inversely proportional to the size of the chunk of video being annotated. Each time a chunk of video is divided into smaller units, I needed to supplement the descriptions of the “mother” chunk to maintain uniqueness. Figure 2 illustrates the different chunks that were needed to rigorously describe the 3,289 contiguous frames which made up the street cleaner sequence. The numbers at the top of each chunk are frame numbers and the number in braces is the size of the chunk. As the size of the chunk of video de-

creased the more detailed the descriptions became. The chunk that consisted of all the footage shot that morning is called a “coarse grained description.” The annotation for the frame with the girl carrying the bananas is a “fine grained description.” It follows however, if I would want to pick out objects within the frame (which a digital environment promises) I would need to find out what was unique about the globs of pixels that make up the objects in the frame. The granularity problem needs to be addressed at the level of the frame in order to describe chunks of video. The granularity problem needs to be addressed at the level of the pixel in order to describe objects within the frame

Figure 2: Various Granularities of Description for Street Cleaner Sequence.

COARSEST	90153-93442 [3289] 11Jul91(am), San Cristobol		
COARSE	90153-92096 [1943] 11Jul91(am) San Cristobol, Fish, Truck	92096-93442 [1346] 11Jul91(am) San Cristobol, Street Cleaner, Barrel	
FINE	92096-92550 [454] 11Jul91(am) San Cristobol Street Cleaner, Barrel, pushing	92551-92875 [324] 11Jul91(am) San Cristobol Street Cleaner, Boy, Barrel, Trash	92876-93432 [556] 11Jul91(am) San Cristobol Street Cleaner, Barrel, Man
FINER		92562-92620 [58] 11Jul91(am) San Cristobol Street Cleaner, Boy, Barrel, Trash, Girl, Banana	
FINEST	92562-92583 [21] 11Jul91(am) San Cristobol Street Cleaner, Boy, Barrel, Trash, Girl, Banana, Entering Screen	92584 (frame) 11Jul91(am) San Cristobol Street Cleaner, Boy, Barrel, Trash, Girl, Banana, Center of Screen	92585-92620 [35] 11Jul91(am) San Cristobol Street Cleaner, Boy, Barrel, Trash, Girl, Banana, Exiting Screen

The more I examined the video, the more I saw. As I continued to analyze what I had shot, my perception of the events depicted became sharper. As a result, I was compelled to create more segments. For example, I became interested in what type of plant the boy was using as a broom; I noted that there was a car driving down the street etc. I had to create a new segment for each of these components. Each



one had to be described independently. In doing so, I had “atomized” the footage into independent chunks.

Another problem arose with the awareness that my descriptions were based on my current state of understanding at the time of recording. Eventually my understanding would change. The in and out points that I used to define a segment during logging would be different from the in and out points that I would eventually use in the editing room. I realized that I would have to go back and redefine the in and out points in regard to how they would be used in a final edited sequence. It was futile to define both in and out points of shots because I could not with any certainty anticipate shot boundaries that would be most useful in a final edit.

### *3.1.2 Strategy Two: Content Markers*

The descriptive overhead of segmentation was too high. And the more time I spent logging the less time I was out shooting. I switched to a more economical method called “Content Markers.” A content marker is defined by one point. Content markers serve as a “quick and dirty” index into the material. It required half the work of segmentation. Content markers allowed me to focus on the quality of description. I could densely annotate the raw footage. While experimenting with this strategy I realized that my free text descriptions served as personal mnemonic markers for the video tape.

The log depicted in figure 3 represents 26,310 contiguous frames (14 minutes) that were recorded in the Municipio of San Juan Chamula, Chiapas while visiting research assistant Carmelino Santiz Ruiz. Carmelino first showed me his medicinal plant garden and then invited me into his house. Upon entering the house I noticed five cases of Pepsi bottles stacked in the corner. As I video-taped the Pepsi cases, he began to pray at the altar in the room. I then followed him into the kitchen where I asked him about the cases of Pepsi bottles.

Figure 3: Video Log of Healing Ceremony with Content Markers

tape07	84793	and this one also
tape07	85163	for burns ?
tape07	85879	fry it in a comal
tape07	86805	grind it like this
tape07	87557	applied like this
tape07	88050	after four days
tape07	88823	name of burn plant?
tape07	89667	Carmelino's house
tape07	90035	interior w bike
tape07	92263	Pepsi bottles
tape07	92957	lighting candles
tape07	93947	lighting first one
tape07	94260	translation 52.23
tape07	94800	translation 52.41
tape07	94847	beginning the prayer
tape07	96720	squatting and praying
tape07	97913	end zoom INRI cross
tape07	99061	This is the kitchen
tape07	99449	thanks to Dr. Berlin
tape07	99819	hay luz tambien
tape07	100615	grinding corn
tape07	103319	drinking Pepsi
tape07	103757	we only have Pepsi here
tape07	104231	close up of Pepsi cola
tape07	107563	everyone drinks Pepsi
tape07	111103	women walking home

The granularity of this log is roughly two annotations per minute. Each entry or record in the database consists of the tape name; the frame number; a free text description. Content markers are not a new way to describe video. Most video tape logs consist of content markers which are scribbled on a piece of paper as the movie maker reviews the raw footage. However, their effectiveness is wholly dependent on the linearity of the medium. For example, when locating “Pepsi bottles” (frame 92263) the video editor uses the shuttle knob on the editing deck to fast forward to that location. In doing so, all the material that has been recorded up to that point appears on the monitor. The editor sees the context that the “Pepsi bottles” are imbedded within. He will see that it is taking place in Chamula; in Carmelino’s garden; now we are in his house; and in a moment he will begin to pray.

However, when we put the video on a random access system, content markers are no longer constrained by the linear properties of video tape. The frames of video become like records in a database that can be sorted and arranged in a variety of ways. When sorted in chronological order the free text descriptions and keywords provide the context for each content marker. For example, “Pepsi bottles” on the diagram above appears in its chronological order between “interior with bike” and “lighting candles.”

But when searching the database for a particular word, “Pepsi”, the list of content markers returned cannot provide the context. In Figure 4 a database search for the words “Pepsi” among all the video annotations does not provide the needed context. With a computerized database of content markers I can locate video -- the items that I am interested in. But this is not what I want, I need to see the surrounding annotations.

Figure 4: Database Search for Word “Pepsi” without Context

tape06	1707	Pepsi or coke
tape07	92263	Pepsi bottles
tape07	103319	drinking Pepsi
tape07	103757	solo hay Pepsi
tape07	104231	close up of Pepsi cola
tape07	107563	everyone drinks Pepsi
tape13	74721	Pepsi bottles
tape14	93083	Pepsi .. Fanta
tape15	28487	Pepsi and orange
tape11	106501	arranging Pepsi
tape23	96843	Pepsi inside Antonio’s house
tape23	108901	Pepsi delivery

Content markers work because they are wedded to the linearity of the medium. But in a random access system the linear integrity of the raw footage is erased. In turn the contextual information that relates to the environment where the video was shot is also destroyed. *Ad hoc* chunks of frames which have nothing to do with the context in which they were filmed. To resurrect the archeological metaphor: the process of making a database of segments or content markers is like the hack archaeologist who shovels the artifacts into a cigar box and then makes inferences about them without considering the site in which they were found. What is required is a method to record this contextual information so that it can be recovered and re-used at a later time.

### 3.1.2.1 Keywords and Context

The content marker descriptions are analogous to the ethnographer’s scratch notes. The effectiveness of content markers is dependent on my memory of the events and utterances that occurred while recording. Their usefulness and significance is not easily transferred to others who were not there. Content markers were a quick and easy way to attach text to the video. Yet, as I recorded more and more hours of video, I needed a more consistent way to peruse the whole set of video tapes.

At this point, I began to use keywords in addition to free text descriptions. Keywords provide a more generalized way to describe video. I used keywords to provide consistent descriptions from one tape to the next. With keywords, I could consistently find related chunks of video among the 27 hours of video tape that I had shot. As shown in Figure 5, the keywords of “Chamula” and “Carmelino” remain

constant while content markers specifically identify what is happening at a given moment. Keywords provide the context for content markers. Now, I could do a database search for Pepsi and the keywords that are attached to that record provide me with important contextual information. But in order to create and retrieve this new type of annotation I had to pay a price. Here, I was faced again with the same problems that arose during the segmentation process -- I had to deal with redundancy in description in a random access system. Sets of generalized keyword descriptors remain constant while the other finely grained descriptions change and evolve. The gains that I had made for not atomizing the footage into independent chunks were wiped out by the overhead of assigning keywords for each content marker.

Figure 5: Content Markers with Keywords Sorted by Frame Number

tape07   84793   and this one also	Carmelino	Chamula	garden
tape07   85163   for burns ?	Carmelino	Chamula	garden
tape07   85879   fry it in a comal	Carmelino	Chamula	garden
tape07   86805   grind it like this	Carmelino	Chamula	garden
tape07   87557   applied like this	Carmelino	Chamula	garden
tape07   88050   after four days	Carmelino	Chamula	garden
tape07   88823   name of burn plant?	Carmelino	Chamula	garden
tape07   89667   Carmelino's house	Carmelino	Chamula	house
tape07   90035   interior w bike	Carmelino	Chamula	house bike
tape07   92263   Pepsi bottles	Carmelino	Chamula	house Pepsi
tape07   92957   lighting candles	Carmelino	Chamula	house praying
tape07   93947   lighting first one	Carmelino	Chamula	house praying
tape07   94260   translation 52.23	<< long text omitted>>		
tape07   94800   translation 52.41	<< long text omitted>>		
tape07   94847   beginning the prayer	Carmelino	Chamula	house praying
tape07   96720   squatting and praying	Carmelino	Chamula	house praying
tape07   97913   end zoom INRI cross	Carmelino	Chamula	house praying
tape07   99061   This is the kitchen	Carmelino	Chamula	house
tape07   99449   thanks to Dr. Berlin	Carmelino	Chamula	house PROCOMITH
tape07   99819   hay luz tambien	Carmelino	Chamula	house
tape07   100615   grinding corn	Carmelino	Chamula	house corn
tape07   103319   drinking Pepsi	Carmelino	Chamula	house Pepsi
tape07   103757   we only have Pepsi	Carmelino	Chamula	house Pepsi
tape07   104231   close up of Pepsi cola	Carmelino	Chamula	house Pepsi
tape07   107563   everyone drinks Pepsi	Carmelino	Chamula	house Pepsi
tape07   111103   women walking home	Carmelino	Chamula	house

### 3.1.3 Strategy Three: Stratification

The implementation of keywords with content markers echoed many of the problems experienced with segmentation. The more detailed a particular chunk of video was annotated the more work was required to describe it using keywords. It becomes evident that descriptions of content have a lot to do with the linearity of a medium. In a random access system we can't rely on the linearity of the medium to provide us with a coherent description. Accordingly we need a new type of descriptive strategy.

The third strategy of logging called “Stratification” represents a shift in the way of creating content annotations for video in a random access database. When sorted on frame number, the content markers became embedded in patterns of keywords. These patterns illustrate the contextual relationships among contiguously recorded video frames. It also illustrates how this context is wedged to the linearity of the medium. We can now trace, in this pattern, what was shot and where. Content markers with keywords produce a layered representation of context. These layers are called strata Figure 6.

Figure 6: Log of Content Markers with Strata.

tape07	85163	for burns ?	Carme	no Cham	la garden
tape07	85879	fray it in a comal	Carme	no Cham	la garden
tape07	86805	grind it like this	Carme	no Cham	la garden
tape07	87557	applied like this	Carme	no Cham	la garden
tape07	88050	after four days	Carme	no Cham	la garden
tape07	88823	name of burn plant?	Carme	no Cham	la garden
tape07	89667	camelinos house	Carme	no Cham	la house
tape07	90035	interior w bike	Carme	no Cham	la house bike
tape07	92263	pepsi bottles	Carme	no Cham	la house pepsi
tape07	92957	lighting candles	Carme	no Cham	la house praying
tape07	93947	lighting first one	Carme	no Cham	la house praying
tape07	94260	translation 52.23	<< long text omitted >>		
tape07	94800	translation 52.41	<< long text omitted >>		
tape07	94847	begining the prayer	Carme	no Cham	la house praying
tape07	96720	squatting and praying	Carme	no Cham	la house praying
tape07	97913	end zoom INRI cross	Carme	no Cham	la house praying
tape07	99061	This is the kitchen	Carme	no Cham	la house
tape07	99449	thanks to dr berlin	Carme	no Cham	la house procomith
tape07	99819	hay luz tambein	Carme	no Cham	la house
tape07	100615	grinding corn	Carme	no Cham	la house corn
tape07	103319	drinking pepsi	Carme	no Cham	la house pepsi
tape07	103757	solo hay pepsi	Carme	no Cham	la house pepsi
tape07	104231	close up of pepsi cola	Carme	no Cham	la house pepsi
tape07	107563	everyone drinks pepsi	Carme	no Cham	la house pepsi
tape07	111103	women walking home	Carme	no Cham	la house

Legend for Strata Lines:

■ Carmelino    ■ Chamula    ■ house    ■ garden    ■ praying    ■ pepsi

We have rotated the method. Instead of creating a database of chunks of frames, we can make a database of descriptions which have frame addresses. These descriptions are called strata. This methodological shift is subtle but critical. The linearity of the medium is preserved with stratification because each description has a temporal/linear extent. With segmentation, the linearity of the medium is destroyed because we have broken up the footage into *ad hoc* chunks. The keywords form strata that capture changes in descriptive state which the camera recorded.

Let us continue with the analogy of the archaeologist. We must analyze an artifact as it relates to its context. When shooting we have the context of where the shoot is occurring: the environmental

context. Also in the environment are the moves that the video maker used to record the scene (camera movements, shot duration, etc.). When recording, a contiguous/linear set of video frames becomes inscribed with this environmental context. Of course, successive shots on a given video tape will share a set of descriptive attributes that result from their proximity. These shared attributes rely on the linear context of the moving image. During recording, the linear context of the frames and the environmental context of the camera coincide. The environmental context is the “where,” “who,” “what,” “when,” “why,” and “how” which relate to the scene; it’s the physical space in which the recording takes place giving it a unique identity. If you know enough about the environment in which you are shooting you can derive a good description of the images that you have captured using stratification. Any frame can have a variable number of strata associated with it or with part of it (pixel). The content for *any* set of frames can be *derived* by examining the union of all the contextual descriptions that are associated with it.

In other words, content can now be broken down into distinct descriptive threads or strata. One stratum constitutes a single descriptive attribute which has been derived from the shooting environment. When these descriptive threads are layered one on top of the other they produce descriptive strata from which inferences about the content of each frame can be derived. Stratification is an *elastic* representation of the content of a video stream because descriptions can be derived for any chunk of video. One only has to examine the strata that any chunk of video is embedded in.

Segmentation (the conventional approach used in computerized video logging systems) forces the user to break down raw footage into segments denoted by begin and end points: such a divide and-conquer method forsakes the whole for the part. Coarser descriptions have to be included at this level of specification in order to describe one frame independently. As we have seen, if a chunk of video as small as an individual frame, its description, in order to be independently retrieved, must encompass larger descriptive units. In segmentation, the granularity of description is inversely proportional to the size of a given chunk of video. The inverse relationship arises out of the need to describe each image unit independently. In contrast, stratification is a method which produces layers of descriptions that can overlap, be contained in, and even encompass a multitude of other descriptions. Each stratum is an important contextual element: the union of several of these attributes produces the meaning or content for that piece of film. Moreover, each additional descriptive layer is automatically situated within the descriptive strata that already exist. The descriptive overhead is reduced. The user can create descriptions which are built upon each other rather than worrying about how to uniquely describe each frame independently. Contiguously recorded shots share a set of descriptive attributes that result from their proximity. These attributes constitute the linear context of the medium. But as we have seen, when loaded on a random access system, the linearity of the medium is compromised. Stratification is a way to maintain the lin-

ear property of the recording medium in a random access system. In this way, rich descriptions of content can be built on top of each other without the redundancy of segmentation.

### 3.2 *Beyond Logging: Video as Resource*

In addition to logging, film makers need tools which will enable them to take segments of raw footage and arrange them to create meaningful sequences. Editing is the process of selecting chunks of footage and sound and rearranging them into a temporal linear sequence (Davenport, Aguierre Smith, Pincever 1990). The edited linear sequence may bear no resemblance to the environmental context that was in effect during recording. During the process of conventional editing, the raw footage that was shot is separated from the final motion picture. In order to create a motion picture with a new meaning which is distinct from the raw footage, a new physical object must be created - an edited print for film, or an edited master for video. It follows that the descriptions of content are directly correlated with objects in the real world. In film, the rushes are cut up into short clips. These short clips are either given a number or a name to be used as a reference during the final production. In video, multiple copies are made. There is the source tape and then the final edited master tape. Each has a distinct name. The correlation of one description to one object is not problematic in these production styles. Often, these objects are not even named -- their status as an independent object is a testament to their existence.

In his thesis, Constraint Based Cinematic Editing (1989), Rubin talks about how the development of new technologies actually encourage different ways of working with the medium<sup>1</sup>. New technologies not only constrain the mode of production and by extension the final form of the final motion picture but also how we think about describing content. The advent of “non-linear” or random access editing systems represents a technological shift where it is possible to *virtually* edit a motion picture -- chunks contiguously recorded images and sound can be called up and assembled on the fly. On these systems there is no need to make a copy of the source material to create a finished work. All that is required is a pointer to the video resource that is available on the system. One shot is played back and then the next shot is retrieved and played back without a gap in between cuts. Movies consist of an edit list which serves to guide the assembly in real time. Virtual video editing is a form of compression. A chunk of video only has to be *referenced* in the movies where it is used. The need to have multiple copies of a chunk of video is obviated because a movie is no longer an actual object but a virtual object.

Editing on a random access video database system is radically different, the link between source and edit does not have to be broken. The shot in the final movie will be the same chunk that is

---

<sup>1</sup>See Rubin’s taxonomy of motion picture production and how different ways of working with the medium affect how the editor organizes the footage into a final motion picture (Rubin, 1989 : 11-15).

logged and annotated as source material. In other words, in random access systems, the distinction between source and final edit becomes so blurred that all video becomes a *resource* which can be used in different contexts (movies). Conceivably, one can edit the source material in the video database into a documentary movie that will be played back on the computer. Moreover, these “edited-versions” can later be used by someone else to make another movie production. Video resources can exist in many different contexts, in many personalized movie sequences. On these systems there will be multiple annotations or names for any given chunk of source material: the first reflects the context of the source material and all others are annotations that are related to a playback time for a personalized movie script. On the computer system there is only one object which is being stored but this object has a multitude of names that are applied to it. Annotations for any chunk of video in the database reflect the context of the source material and also the context of how it is assembled into a sequence. The video resource can be two things at the same time - the initial descriptions reflect the context of where it was generated while the play list reflects the context of a sequence. The raw footage needs to be described along with how the footage is used in the different movies.

When editing a sequence, important relationships in the raw footage come to the fore. In an edited sequence causal and temporal relationships in the raw footage are made explicit. The content of the material dynamically changes from the environment where it was shot to its location in an edited sequence. Since cinematic content is inextricably linked to context, *the significance of the source material becomes transformed through use*. The Anthropologist’s Video notebook requires a computer representation for the content of video that dynamically changes with the context. Stratification will provide a computational environment where chunks of video are annotated in terms of the context of where they were created but also in terms of the context of where they are re-used in different movies. Thick descriptions of content of a video stream’s content are dynamically built over time as they are used and re-used.



#### 4. REPRESENTATION AND DESCRIPTION IN MOTION PICTURE PRODUCTION

In the last chapter stratification was derived in the context of a real research setting. The moves that led to its development were pragmatic. In order for stratification to be the basis of an entire video database system must support content annotations that reflect the context of where the video stream was recorded and also how the video stream is re-purposed during sequence assembly.

Consider some of the implications of the moves that went into the development of the stratification method. First, the video ethnographer had an intention to study something - he was at a particular place and at a particular time. Not only was he present, he interpreted what was unfolding before him and decided when to record video and how the images were composed in the frame. In this light, the video tape is a trace or a description of the event that was witnessed in terms of the anthropologist's intention of being there and his interpretation of what happened while shooting the video. The images that are recorded on a video tape are artifacts in the sense that they are *audio visual descriptions* that reflect the maker's interpretations and intentions while observing an event. Moreover, when annotating the video stream in a video database system another type of *textual description* is applied. These textual descriptions reflect the researcher's memory of the context of what was going on during the shoot. In a video database application *lexical* descriptions are applied to *audio visual* descriptions of an observed event.

If video becomes a resource on a random access system, what then are the implications for the design of an information system that can support the type of research activity described above? The Anthropologist's Video Notebook must enable researchers to manage two different types of descriptions that are derived from two different media forms - text and video. We need a digital representation for lexical and visual descriptions that is integrated in the same system. Moreover, the computer representation for video ethnography must support dynamically changing content and the generation of new content. It has to be generative in order for a human or a machine to find appropriate footage and then arrange this footage in a coherent sequence.

Now it should be pointed out that for an interactive video application such as the Anthropologist's Video Notebook we need to have a computer representation for the medium. Furthermore this representation must allow us to describe our memories and observations in a way that we can communicate them to other people. The terms representation and description are not synonyms. A clear understanding of these terms is a requisite for explaining the complex process of how a medium is transformed during ethnographic video production. The distinction between representation and description is well developed in the field of Visual Cognition where the vision is conceived as an informa-

tion processing task. Although David Marr in his book *Vision* (1982) is interested in the processing and representation of visual information, his definitions will serve our purposes. He writes, "Vision is a process that produces from images of the external world a description that is useful for the viewer and not cluttered with irrelevant information" (Marr, 1982 : 31). Although, Marr does not specifically focus on image recognition or the naming of images, he does explore the *process* of how people extract and use information from images. Marr is interested in the process of how people extract a useful description of the world from the images that they see. He explains:

The study of vision must therefore include not only the study of how to extract from images the various aspects of the world that are useful to us, but also an inquiry into the nature of the internal representations by which we capture this information and thus make it available as basis for decision about our thoughts and actions (Marr, 1982 : 3).

Marr strives to understand how people capture information from images in terms of internal mental representations. A computational theory for vision requires that we look at the process as an information processing task.. Marr is committed to the idea that there is something in an image that is reliably captured, at least in a computational sense, by vision. Does the mind pick out a useful description of objects in the world? Marr realizes that "in order to understand a device that performs an information processing task, one needs many different kinds of explanation" (Marr, 1982 : 4).

An information processing task involves three components namely, the process and the inputs and the outputs. Although this I/o model or "black box" approach is somewhat simplistic, it can serve to clear up the confusion. Marr defines the terms "representation" and "description" in regard to information processing tasks. He makes the following distinction: "a representation is a formal *system* for making explicit certain entities or types of information, together with a specification of how the system does this." While a description is "the *result* of using a representation to describe a given entity in that representation" (Marr, 1982 : 20 (emphasis added)). In other words, a representation consists of a formal scheme or *set of rules* for putting symbols together. A description is the output of a representational scheme.

Marr is concerned about the usefulness of descriptions and how they are dependent upon representational systems. The choice of a representation impinges upon what someone can do with a particular description of an object:

The notion that one can capture some aspect of reality by making a description of it using a symbol and that to do so can be useful seems to me a fascinating and powerful idea. But even the simple examples we have discussed introduce some rather general and important issues that arise whenever one chooses to use one particular representation. For example, if one

chooses the Arabic numeral representation, it is easy to discover whether a number is a power of 10 but difficult to discover whether it is a power of 2. If one chooses the binary representation, the situation is reversed. Thus, there is a trade-off; any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover.

This issue is important, because how information is represented can greatly affect how easy it is to do different things with it. This is evident even from our numbers example: It is easy to add, to subtract, and even to multiply if the Arabic or binary representations are used, but it is not at all easy to do these things—especially multiplication—with Roman numerals. This is a key reason why the Roman culture failed to develop mathematics in the way the earlier Arabic cultures had (Marr, 1982 : 21).

There is a trade off when choosing a representational system. The choice of one representation system over another is a commitment to a specific type of description.

#### *4.1 An Image is a Description*

We can now view the practice of video ethnography as an information processing task and think of the camera as a representational system. The picture that the camera takes is a description of some object in the real world in terms of the current state of the camera. If a photographer changes the state of the camera by resetting the focal length of the lens or by using black and white film instead of color then the resulting photo will correspondingly change. The photographic image is a description of some object in the world. The camera is a representational system that mediates the object by producing descriptions. Many different types of images can be created of the same object. An image considered outside of the context of the representational system where it was created is potentially ambiguous — many photographs can correspond to the same object.

The symbolic mapping of an object to a description pushes other types of information into the background. This idea is demonstrated by Wittgenstein in the following example:

...if water boils in a pot, steam comes out of the pot and also pictured steam comes out of the pictured pot. But what if we insisted on saying that there must also be something boiling in the picture of the pot (Wittgenstein, 1958 : §297).

Before attempting to describe the content of the picture of the teapot, three interrelated elements need to be considered. First there is the “real” teapot with water boiling in it. Then there is the representation system that maps from the real teapot to the picture. And the final element is the picture of the

teapot “the description”. It doesn’t make sense to think that there is water boiling in the pictured pot if we take into account that the picture is an artifact of some representational system.

An image is an interesting type of object because a multitude of representation systems can produce one: the human eye, the camera, etc. In light of Marr’s distinction between representation and description - the image is clearly a description that is actually re-mapped through a multitude of representational systems.

#### *4.2 The Descriptive Dimensions of a Video Stream*

There are two types of descriptions which operate on a video stream. One has to do with the medium as a representational system --the design space where content can be generated. The other refers to the way that knowledge functions as a representation system which produces consistent lexical descriptions of content. On the one hand, the video stream enables us to produce descriptions of actions and utterances which are objects - sets of contiguously recorded frames. And on the other, we have knowledge about the content of the video stream which are lexical descriptions.

The representation of the image must allow different types of understanding to operate for the same set of moving images. We have two different design environments one for motion picture production and one for the formation of lexical descriptions of motion pictures. The design environment for movie production and the design environment for the generation of knowledgeable descriptions are temporally linked in a random access video database system. Stratification provides a framework where the video image is alive - where the different ways to describe a video stream are linked in time. Here a critical distinction must be made-- before we can represent knowledge about an image -- first, we must develop a representation of an image itself. This distinction is subtle but critical. We do not want to create a system that decrees the truth or the facts about video. Instead, we want a system that allows images to be contested by different people with a multitude of intentions. Stratification is a representation of the video stream that can support a multitude of descriptions that are in turn the result of various visual and lexical representation schemes.

In light of the above implications the moving image database exists at the juncture of the design environment and the naming environment. Scratch notes that were generated in the field and served as memory aids become incorporated into thick description. These thick descriptions then guide makers to video content which then can be re-used to make movie sequences. The sequence annotations can also be used to generate new content. The process is recursive as video resource iterates in the design environment.

## 5. AUDIO VISUAL DESCRIPTIONS BY DESIGN

The video ethnographer interacts, interprets and manipulates two different media types to produce content. In making a video database system, the way that the creative process behind the construction of content is understood directly influences the type of content that can be generated by the system. It follows that if we want to represent content then we have to create a computational space where motion picture content can be generated and *designed*. The computational space must be an *environment* that will support how video makers use their knowledge of working and interacting with a medium to make content.

Exploration and experimentation are key features of the creative process but they are also the hardest elements for the makers themselves to articulate and understand. Nevertheless, it is exactly this type of knowledge that will be the foundation of a computer representation of the moving image. How then can we make this knowledge explicit?

The medium of video constrains the maker in terms of the possible moves. In spite of these constraints new content can be generated. A *design environment* can be defined as a medium together with the set of moves that can be made on that medium to generate content. The successful implementation of a moving image database that can be used by movie makers or automatic story generators depends on understanding the process of how the maker produces content by interacting with the medium in a design space. According to Schön some of these approaches have not framed the problem properly. He explains:

Some of the best minds engaged in research on design computation have focused on the problems of developing computational representations of design knowledge — in effect on the problem of building machines that design. When we think of designing as a conversation with the materials conducted in the medium of drawing and crucially dependent on seeing, we are bound to attend to processes that computers are unable — at least presently unable — to reproduce: The perception of figures or gestalts, the appreciation of qualities, the recognition of unintended consequences of moves.

It does not follow from this that computers can have no significant use as design assistants. What is suggested, on the contrary, is that research should focus on *computer environments that enhance the designer's capacity to capture, store, manipulate, manage and reflect on what he sees* (Schön & Wiggins, 1988 : 31 (emphasis added)).

Schön's suggestion is an interesting way to frame the problem. Although it is difficult to get insight into designers internal processes that inform particular design decisions we can create computa-

tional environments that enhance and facilitate the expression of a designer's *intentions*. A design environment can be conceptualized as a space ( which can have cognitive, discursive and physical elements) where a particular set of design moves can be applied to a medium to produce content. In this way, a designer's work is constrained by the types of manipulations that are possible for a given medium and by the environment where content is created.

### *5.1 Toward a Theory of Video as Design*

It is difficult to articulate the subtext of movie making as a process. Paradoxically, we need to understand this process to create a computational environment for the creation of motion pictures. When discussing how he edited "Les Oeufs a la Cocque" Richard Leacock explains that he would first select a whole event or action from the hours of raw source video that was shot. And then, after viewing this selected footage repeatedly -- to the point of almost memorizing it he would leave the editing room. He would visualize how the final sequence would be fitted together while doing some other type of activity -- for Leacock the design process can occur in the background<sup>1</sup>. He brings a vast amount of knowledge into the editing process -- his work is a testament to the efficacy of this process.

Each motion picture design environment has distinct properties that affect the types of movies that can be made. The development of new technologies leads to new forms of cinematic story telling (Davenport, Aguiere Smith, & Pincever, 1991). Nevertheless, film makers are constrained by technology. Leacock underscores this point. He talks about how motion picture technology has historically constrained the creative process and as a result what could be expressed in the medium:

In 1959, Jean Renior described the problem of motion pictures technology in an interview with film critic Andre Bazin and Roberto Rossellini: "... In the cinema at present the camera has become a sort of god. You have a camera, fixed on its tripod or crane, which is just like a heathen altar; about it are the high priests -- the director, cameraman, assistants -- who bring victims before the camera, like burnt offerings, and cast them to the flames. And the camera is there, immobile -- or almost so -- and when it does move it follows patterns ordained but the high priests, not by the victims" (Leacock, 1986).

Since that time, Leacock explains:

Progress has been made. ... For me personally, the greatest change took place three years ago when I retired after twenty years of teaching at MIT. I joined the 20th century by acquiring a computer and going to work on Video-

---

<sup>1</sup>Personal communication January 1992.

8. Ever since the introduction of the CCD and the refinement of video editing equipment I have come to love to video image. ... For the first time we can all work as Flaherty liked to work. Shooting what ever we choose, as much as we choose, when we choose. Editing at home so that we can go over and over our material, shoot some more, come back to look, edit again ... and again. Try new approaches, experiment with lenses, with ways of shooting, with the relationships to the people we are depicting, to the people we work with. Create different version for different situations. There is no limit ... (Leacock, 1991 : 9).

Although a motion picture is made under many constraints (technological, financial, etc.), the movie maker works within these constraints through experimentation and reflection. The design environment involves the interplay of many interrelated elements such as the film maker's intentions, the media, and the possible design moves given a particular production environment. These elements contribute in some unpredictable ways to the development of content. Leacock writes about the experience of experimenting during the process of shooting:

As an apprentice cameraman totally immersed in the agony as well as the ecstasy of making a film, I learned to look through the camera, to search, to pry, to experiment and then to watch -- as one accused -- the rushes. ... I mean that hard -to -define and rarely- found quality of there being a love affair between the film maker and the image ... By "love affair" I do not imply that you have to love what you are filming; in fact, you may hate it but you are involved emotionally, or intellectually. You are engaged. What you are doing is not just a job that ends when you get paid or at five o'clock. For me there must be pleasure. I do what I do for the pleasure that is involved. I may be tired by circumstances. I may have terrible time getting what it is that I am after. I may not know what I am after, but when I get it I know and it gives me tremendous satisfaction (Leacock, 1986).

Leacock talks about total immersion in the medium. He might not know what he wants to get but he will know as soon as he gets it. Through the process of searching and experimenting he is successful. But in what ways can we think of this type of exploratory practice as rigorous experimentation -- let alone create a computer model for this type of activity?

Donald Schön finds a similar problem with architects when they try to explain their individual processes of design. Unfortunately it is very difficult to make such explanations of creating and designing into computational theory:

Designers are usually unable to say what they know, to put their special skills and understandings into words. On the rare occasions when they try to do so, their descriptions tend to be partial and mistaken: myths rather than accurate accounts of practice. Yet their actual designing seems to reveal a

great deal of intelligence. How, then, if we reserve 'knowledge' for what can be made explicit, are we to explain what designers know? If, on the contrary, we recognize designers' *tacit* knowledge, what shall we say about the way in which the *hold* it, or get access to it when they need it? (Schön, 1988 : 181).

Although the movie maker's knowledge is tacit it is always wedded to the medium. What we need to understand is how the makers interact with the medium of moving images to create a finished work. Schön has written about ways to think about the design knowledge of architects. His conceptualization of the design process will serve as a useful model for the development of a theory of motion picture production as design. For him, "designing is not primarily as a form of 'problem solving', 'information processing', or 'search', but as a kind of *making* ... design knowledge and reasoning are expressed in designer's transactions with materials, artifacts made, conditions under which they are made, and manner of making" (Schön, 1988 : 182). The design process is intimately related to a particular maker's transaction in a medium.

The production of a motion picture involves both shooting and editing. These two activities occur in two distinct design environments. The types of manipulations or moves that are possible in each environment are constrained by the malleability of medium and the technology used to create the motion picture. The intentions of the movie maker are played within these constraints to produce content. Content is never fixed during the design process but it evolves in each design environment.

At this point it will be useful to formalize some concepts for talking about the design process. Schön characterizes the process of a designer interacting with the medium as "reflection-in-action" (Schön, 1987) or "conversational learning":

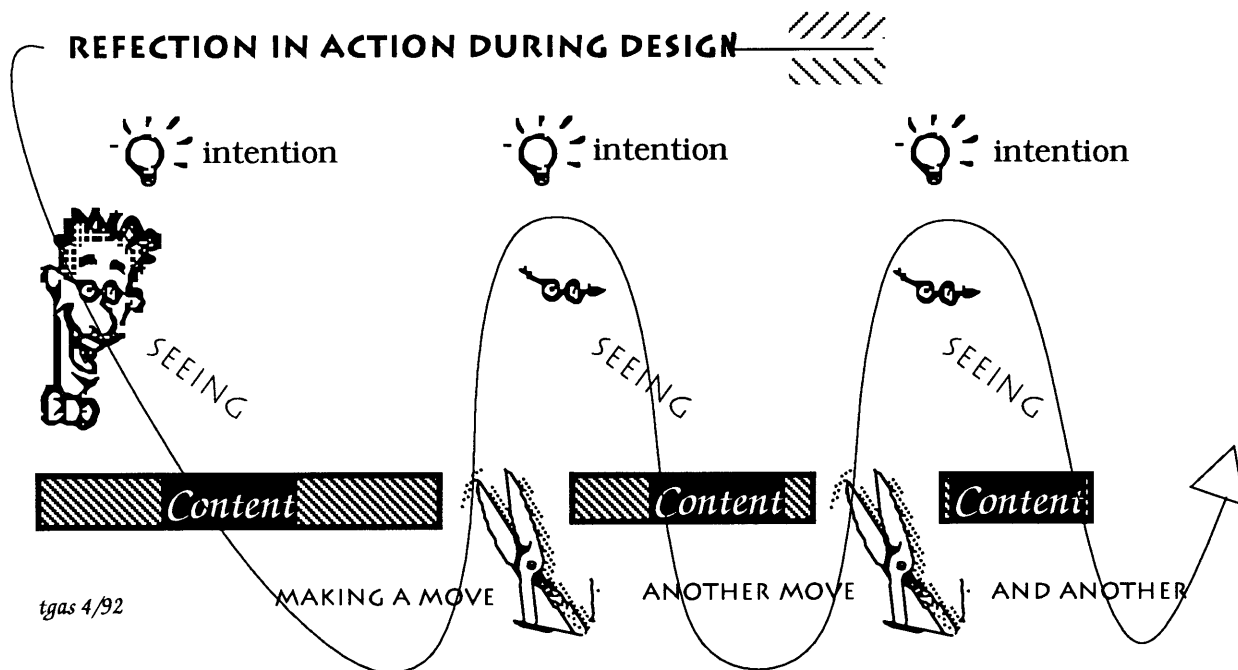
By this we mean the gradual evolution of making something through reflective "conversation" between makers and their materials in the course of shaping meaning and coherence. "Reflective" has at least two intentions here and often they are intertwined as to be indistinguishable: the makers' spontaneous (and active) reflection response to their actions on the materials, and the "reflection" of the materials (as they take various shapes and forms) back to the makers. The two kinds of reflection can be thought of as two kinds of "talking back". In the first, the makers talk back to the materials (re-shaping them), in the second the materials talk back to the makers, re-shaping what they know about them (Bamberger & Schön, 1982 : 6).

In other words, content is produced when a particular design intention is acted out on the medium. This is a reflexive process. For Schön, design is an evolution of an intention that arises out of the apprehension of qualities. Intention is a vector which points to the future and that arises out of how one sees an object. The designer apprehends the qualities of the current state of the medium -- makes



an evaluation of the current state of content and uses this information to formulate a new intention. This new intention provides the motivation to change the current state of content in the environment. A move is carried out to change the current configuration of content to the next stage. New content emerges out of the previous state of affairs. Reflection in action is a process that involves seeing a certain configuration of the materials and then formulating an intention about them and then making a move which results in the transformation of content. Reflection in action iterates throughout the design process (figure 7). It is like an arrow that pulls both the materials and the designer's intentions up to the final form of content.

Figure 7: Reflection in Action. Through an iterative process of seeing, reflecting and acting the designer creates content. The designer has an intention to create something new. He looks at the material and then makes a move to manipulate it / change it into something new.



## 5.2 Experimenting in a Design Space

The designer experiments by reflecting and acting on a medium. Schön groups this type of experimentation into three classes. The first style of experimentation is hypothesis testing:

Hypothesis testing follows a process of elimination. The experimenter tries to produce conditions that disconfirm each of the competing hypotheses, showing that the conditions that would follow from each of these are not observed. The hypotheses that most successfully resists refutation is the one the experimenter accepts — only tentatively, however, because some other factor, as yet undiscovered, may turn out to be the actual cause. ... In order to stage

such a competition of hypotheses, the experimenter must be able to achieve selective variation of the factor named by competing hypotheses, and must be able to isolate the experimental situation from confounding changes in the environment. ... And to this end, he is expected to preserve his distance from experimental phenomena, keeping his biases from the object of study (Schön, 1987 : 69 - 70).

Although hypothesis testing is the conventional approach to experimenting, Schön argues that it does not adequately capture a key feature of the design process: change. The practicing designer is usually unable to shield his experiments from confounding changes in the environment. In fact, "the practitioner has an interest in transforming the situation from what it is to something he likes better. He also has an interest in understanding the situation, but it is in the service of his interest in change." Sometimes the maker does not have a design intention. An intent is usually found and developed by working with the medium. This type of activity is a second type of experiment.

When action is undertaken *only* to see what follows without accompanying predictions or expectations, I call it *exploratory*. ... Exploratory experiment is the probing, playful activity by which we get a feel for things. It succeeds when it leads to the discovery of something there (Schön, 1987 : 71).

A third type of experiment is the move-testing experiment which involves a cyclical process of seeing - moving - seeing. There are two ways of seeing:<sup>1</sup> seeing a pattern and seeing a quality. The way that the designer sees a particular configuration of materials (or problem) formulates a design intent. A move is then a shift in the configuration of the problem.

In the simple case where there are no unintended outcomes and one either gets the intended consequence or does not, I shall say the move is affirmed when it produces what is intended and is negated when it does not. In more complicated cases, however, moves produce effects beyond those intended. One can get very good things without intending them, and very bad things may accompany the achievement of intended results. Here the test of affirmation of a move is not only Do you get what you intend? but Do you like what you get? ... A better description of the logic of move testing experiments is this: Do you like what you get from the action, taking its consequences as a whole? If you do, then the move is affirmed. If you do not, it is negated (Schön, 1987 : 70 - 72).

The move that the designer makes is based on what he sees in order to realize a particular intention. It is the intention of the maker that formulates the problem and tests it. Given any move, the

---

<sup>1</sup> Lecture 1 October 1991 Design Research Seminar (4.278j).

designer either gets what he intended (he liked what he got) or he gets an unintended result that he likes (a great move). The indeterminacy about judgments inherent in each move has an element of exploration. In terms of computational theory, indeterminacy, interpretation, and exploration are some of the hardest things to represent.

### 5.3 *Locating Content in a Design Space*

Content is in a constant state of transition during the design process. Content is dynamic. Schön and Bamberger (Bamberger & Schön, 1982) write about how they discovered how participants in a design exercise produced “transitional objects,” which were configurations of five Montessori bells arranged on a table which served to hold the current “state” of their musical compositions. They explain:

... while the goal of the participants is to make a tune, the evolution towards this goal included making a number of “transitional objects” -- namely a series of constructed and reconstructed bell-arrangements on the table. These transitional objects in their various transformations serve to “hold still” the meanings the participants give to the bells. Each arrangement becomes a reference to these meanings. Each transitional object becomes a *reference entity* an embodied and enacted description of what the participants know so far (Bamberger & Schön, 1982 : 8).

Reference entities are transitional objects that show a path that leads from an intention to a finished product. *A reference entity is a transitional type of content.* “A reference entity serves as a naming function but it does not literally name. That is, a reference entity is unique, often transient and it is “held” by the materials used for making things with the domain” (Bamberger & Schön, 1982: 8-9). The development of reference entities are an integral part of the design process because they allow the designer to reflect on the current state of the work and formulate another design intention to go on to the next step. But most interestingly,

At the same time the (the participants) come to see *these* materials in new ways they are building a *unique* coherence. Unexpected insight *evolves* in the work of making but makers tend only to see it when, through *the evolutionary process of making*, itself they can recognize it. And when they do, the transitional objects, the moves on the way seem to disappear. Practicing a kind of “historical revisionism”, they attribute insight to the moment when it occurs, even finding in the moment a sense of certainty -- of course, “we knew it all the time!” (Bamberger & Schön, 1982: 15 (emphasis in original)).

Although the theory of reflection in action produces many transitional objects along the way these objects tend to be wiped out by the final work.

A finished product -- a computer program that works, a proof that matches a canonical one -- tends to “wipe out” in its clarity and logic especially when expressed in conventional symbolic notations, the conversations with the materials through which they evolved (Bamberger & Schön, 1982: 16).

The ‘we knew it all the time’ evaluation of a design process is perhaps the most difficult aspect to come to grips with when developing a computational environment to support video ethnography. It is as if the medium is the glue for content. When considered in isolation there is nothing attached. The process of reflecting and interpreting the medium within a constrained design space is what delimits content.

The content that is created during the process of reflection in action is indeterminate and under specified because the current state of content is subject to new types of seeing and as a result is prone to moves which transform it to still other types of content. The notion that content is under specified during the production process has important implications for the development of a computerized database of the descriptions of the content of a video stream. Descriptions of content are contingent on a particular design state and depend on the intent of the person entering the descriptions in the database. How can we represent content that is in the process of being transformed? We have to locate content in the design environment in order to describe it.

By operationalizing Schön’s framework for understanding the design knowledge of architects we can identify some key elements for a theory of video ethnography as design. First, video production takes place in design environments where makers can operate on the medium and develop intentions by reflecting and then acting. The way that the transitional content elements or “reference entities” appear and vanish along the way provide essential information that also needs to be represented. Second, each type of design environment constrains the designer both in terms of the development of intentions and the types of moves and manipulations that are possible with the medium. Third, the process of design is dynamic. Content emerges at various moments during the process and is sometimes wiped out by the power of the final work.

Content is produced in two different types of design environments each having its own set of constraints. Content first emerges in the *Design Environment for Shooting* and then is transformed into the final motion picture in the *Design Environment for Editing*. In order to design a video database such as the Anthropologist’s Video Notebook we must first develop a conceptual framework for each of these design environments. Most important, by understanding the constraints of each we can create a robust computer representation of the content of a video stream. In the next section, we will ex-

amine how stratification can be used to computationally represent the process of how the video stream becomes transformed in the motion picture design environments for shooting and editing .

#### *5.4 A Design Environment for Shooting Video*

The motion picture design environment for shooting contains many interrelated contextual elements such as the video maker's intentions, the media, and a set of moves. The video camera is the center piece of the design environment for shooting. The video camera serves as a representation system that allows light and sound waves to be captured on a medium. The images and sounds that a camera records are descriptions of events and utterances that occur in the shooting environment. A moving image is more than just a spatial and temporal sampling of an event or utterance, it is also the result of the video maker's intentions. The video maker's intentions, the current state of the camera (focal length, filters, type of media) and some event happening in the environment coalesce when video tape is being recorded.

First, there is an intention for wanting to make a video. This initial intention is related a video project's goal and the logistics of getting a camera and crew into a place where the action is going to take place. Another set of intentions comes into play while shooting. These intentions involve choices of composition of action within the video camera's view finder and choices of when to begin and end recording a particular shot. The shots of video that have been recorded on the tape are "artifacts" of both the macro intentions that guide the project and the micro intentions of the person behind the camera. These intentions constantly change during shooting. Leacock explains,

I think that filming is a continuous process of learning and it is part and parcel of editing, that the one skill feeds off the other. So many of the decisions made during filming are part of the final edit, and only the people who made those decisions know which work out and which did not. Often, in observational shooting, you will try a number of different ways of capturing an event or object. These different approaches can easily be confused and your whole intent destroyed (Leacock, 1986).

The images and sounds that are recorded on the video tape reflect the film maker's future directed intentions about what the final video tape is going to be like. As they are recorded the motion pictures and sounds are recorded and pulled along an intention vector which points to the final work. The content of the motion picture is the result of a series of moves that were motivated by the maker's intention and constrained by the type of media, the environment where the film was shot and the type of moves that were possible given the technology. Content is the result of developing intentions about what the final product will look like and then putting these intentions into practice by reflecting and acting them out on a medium.

The next important element is the environment itself - the time and place where people, events and actions that are going to be recorded. The environment can be thought of as the “where,” “who,” “what,” “when,” “why,” and “how” - the contextual factors present during recording.

The media is of course another critical element in the shooting environment. The camera produces a contiguous set of frames that are recorded on the medium. The type of media (film, video or digital) determine not only the look of the images that are recorded but also how the images can be manipulated in the editing stage. Other elements are the moves that result the camera recording the images on the medium. By moves, I refer camera movements (crane, dolly, tilt, pan, steady cam etc.); duration of shot; focal length (wide, medium, close-up, zoom); sound recording (sync, wild, no sound); lighting and image quality (control over brightness, contrast, color by the use of filters, gels, lenses, shutter speed, f-stop etc.).

The recorded moving images are descriptions which reflect the above mentioned elements. The intents, the medium (video) and the events that shift and move in time are the context for the design environment for shooting. In this context, the content of the video is produced.

#### *5.4.1 Two Types of Context*

Once recorded, these interrelated contextual elements of the design environment for shooting become artifacts which are physically linked to the medium. The recorded video tape is a detailed record of how a particular video maker made sense out of a situation as it was unfolding. While the images recorded on cassettes of video tape are a physical record of what occurred, the video maker’s memory of the chronology of events, as well as his feelings and decisions that went into a particular shot serve as an ephemeral index into this material. The contextual factors in the motion picture design environment for shooting (the maker’s intentions, the type of media, the moves that were made and environmental factors such as the “where,” “who,” “what,” “when,” “why,” and “how”) are clues about the content of the moving image.

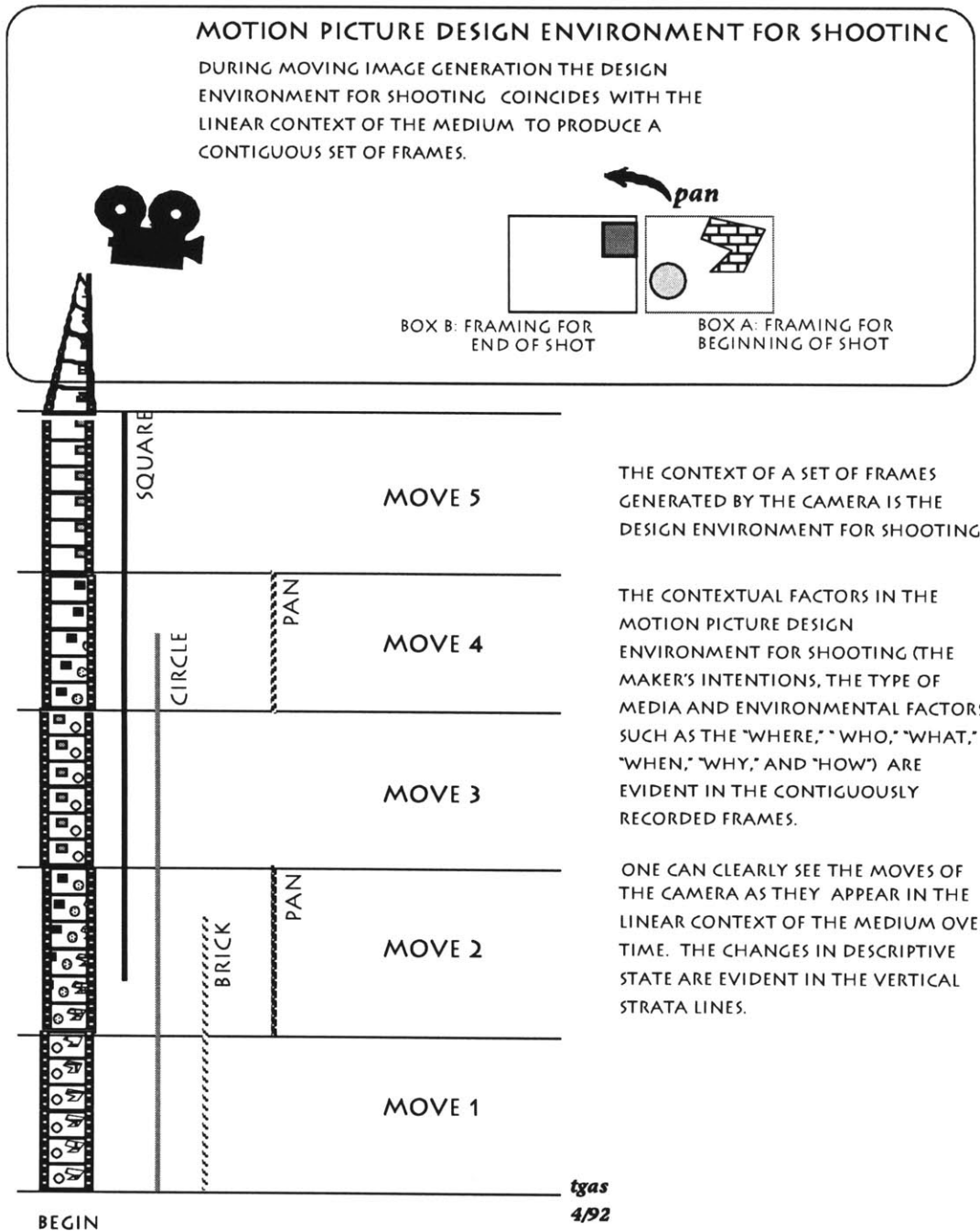
A strong contextual component for the moving image is inherently linked to the medium itself. A motion picture with sound can be described as a series of frames that are recorded and played back, one after another, at a certain rate. In other words, linearity is an inherent characteristic of the medium. The order in which frames are placed is critical to editing. And, as we shall see, adjacency is equally critical for the development of a computer representation of the moving image.

In figure 8 we can examine how two contexts interact when the moving image is being captured. The frames are generated linearly by the camera. Thus, they share a set of descriptive attributes

resulting from their proximity to each other at the time of recording. *The context of a set of frames being generated by a camera is the design environment of the camera.* During moving image generation, the linear context of the frames and the context of the design environment for shooting coincide. Environmental factors which relate to the situation of the camera are inherited by all the frames of the moving image being recorded. We now have a contiguous linear sequence of frames that also shares these environmental factors. The images on the video tape are a stream of visual descriptions of the contextual factors in the shooting environment at a given moment. The strata lines are a synchronous stream of lexical descriptions of those same contextual factors.

The content of a video stream is dynamic. The content changes in time at a rate of the frames that the camera records per second. The five moves (two pans and three steady shots) are clearly visible on the strip of film. The strata lines show how the content of the video strip dynamically change. For a given frame we can get a lexical description of content by examining the strata that the frames are embedded in. The strata lines allow the user to visually browse the lexical description of dynamically changing content.

Figure 8:



In the design space for shooting we have three elements: a square, a circle and a chunk of bricks. These objects have a fixed spatial and temporal relationship to each other - they are at the same place at the same time. The camera is also present in the design space. In the diagram, the movie maker begins recording with the chunk of brick and circle in the frame (box A). The camera then pans



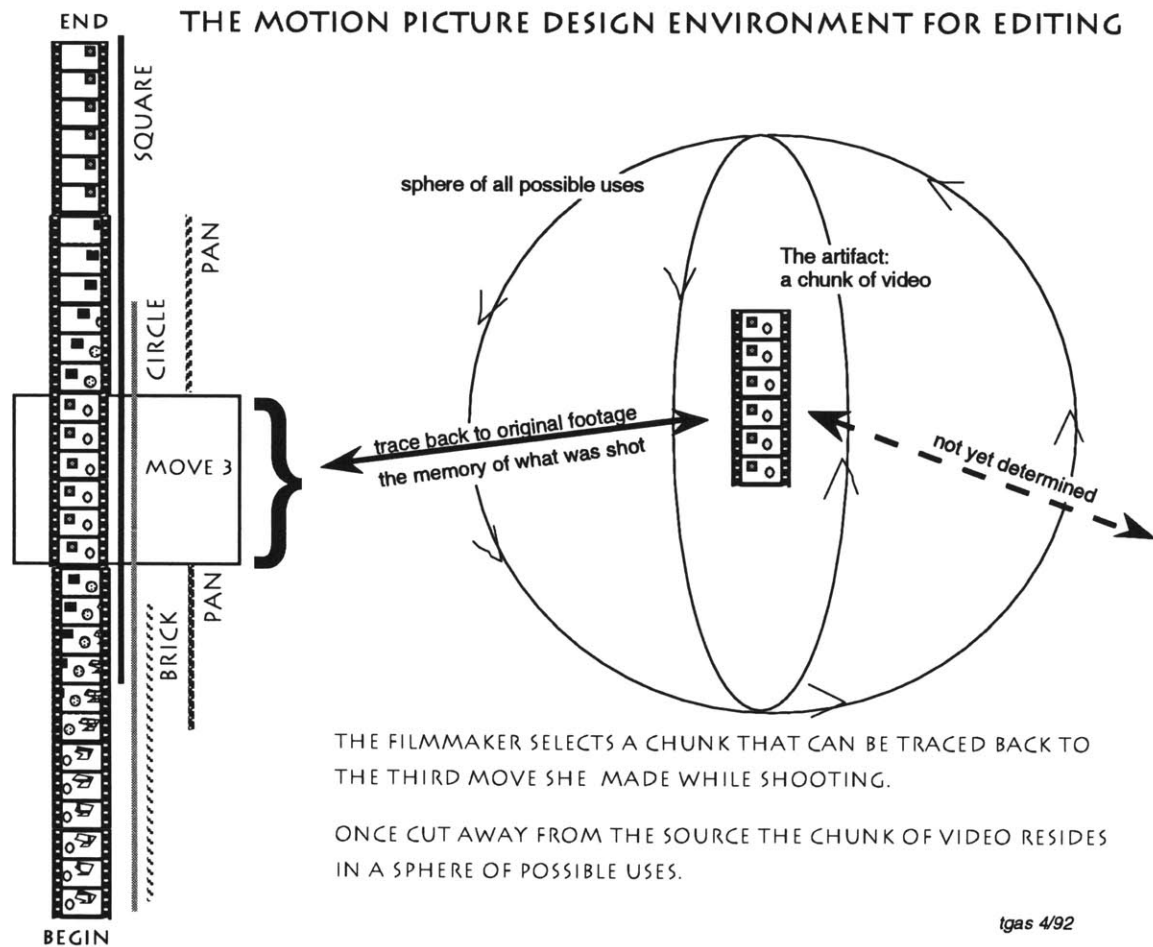
to the square. This final framing is represented with box B. As the camera records the film maker stops for a moment framing the circle and the square - he has made a move based upon some intention - perhaps he liked this particular composition and then later panned until only the square appears in the frame. Although the spatial and temporal relationship between the objects in the shooting design environment remained the same - the intentions of the film maker in using the medium of film produced motion picture artifacts with differing spatial temporal arrangements. In the strip of film which drops vertically from the camera we can see how the contiguously recorded frames are a *trace* of the film maker's moves which are the result of the intentions he developed during the shoot.

When thought of in this way, the motion pictures and sounds that are recorded on the video tape are a trace which points back to the set of intentions of the maker, the environment where the video was shot and also the moves that were elemental to its production. All of this information is contained (or better yet) confounded in the recorded video tape. When viewing an unmarked video tape, the process of recovering this information is analogous to an archaeologist trying to find out what a newly discovered artifact represents. With stratification we can begin to attach lexical descriptions that dynamically change with the video stream. The dynamically changing stream of lexical descriptions which mimics how the images in the frames change over time gives the video stream a sense of itself.

#### *5.4 Design Environment for Editing Motion Pictures*

Of course, the video editor can create a new context by manipulating the linear sequence of image artifacts. In this way the process of editing is a gateway to future intentions that were not manifest in the raw footage. In traditional types of video editing, the film maker organizes pieces of video by cutting or by copying them onto another video tape. From these pieces, the final movie emerges. These chunks of video are transitional types of content in the sense that they reflect the editors intention to use the chunk of film in a particular context. Simultaneously, other chunks of film are removed from their original context -- that is, the original linear context of how they were contiguously shot. These various chunks are reflected on by the maker and intentions are formulated about their eventual arrangement into the final narrative piece. In this way editing is a kind of experimenting. Editors make hypotheses about how the final movie will turn out and explore how different orderings and juxtapositions of sound and images come together to produce the content of the final movie. Figure 9 is a illustration of how a shot is selected by an editor and how this shot relates back to the raw footage. The selected chunk is in a free space -- it can be used in myriad contexts.

Figure 9:



In the diagram above we have the strip of contiguously recorded images that were created in the design environment for shooting. Here in the editing room the movie maker develops an intention which guides the composition of the final movie. The contiguously recorded frames are a temporal slice or sampling through the shooting environment. The frames reflect the moves, intentions and the ambient reality of the shoot. Additionally, there is the memory of the maker, he knows what images came before and after this chunk in the contiguously recorded raw source material. Although the strata lines reflect a lexical description of the recorded content, these descriptions can serve as hooks that allow the maker to locate desired chunks of video so that they can be reused in a new sequence. The tension between the past and future intents is the source for much creative energy in the design environment for editing.

The movie maker uses his memory from the design environment of shooting along with the stratified descriptions of content to locate a chunk. His memory together with the lexical description of

content guide him to the desired piece of raw footage: he wants a chunk that contains a static shot of circle and a square. Next, he makes a move and cuts out the chunk<sup>1</sup>.

Now out of its original linear context (which is a trace of the conditions under which it was generated), the chunk of video enters a “sphere of possible uses” where a host of other intentions can be applied to it. At this moment, the designer’s future directed intention starts to take hold. The chunk of video which resides in the sphere of possible uses is the result of his past intention and also the object of new intentions. The intention shifts from the shooting environment where the leading question was “How do I record an event that is in the process of unfolding?” to the editing environment where the leading question becomes “How can I order or structure moving images artifacts with the purpose of communicating something to an outsider viewer?”

The constraints of the design environment - the moves which are possible for a given medium - are reflected in how the designer develops an intent. The video maker satisfies his design intents by ordering chunks of motion pictures into sequences with the use of optical effects, (fades, dissolves, wipes); use of sound (ambient, live, voice, effects) etc. If the chunk turns out to be too short - he can go back and get the extra frames. If the adjacent frames do not satisfy these requirements other moves have to be made.

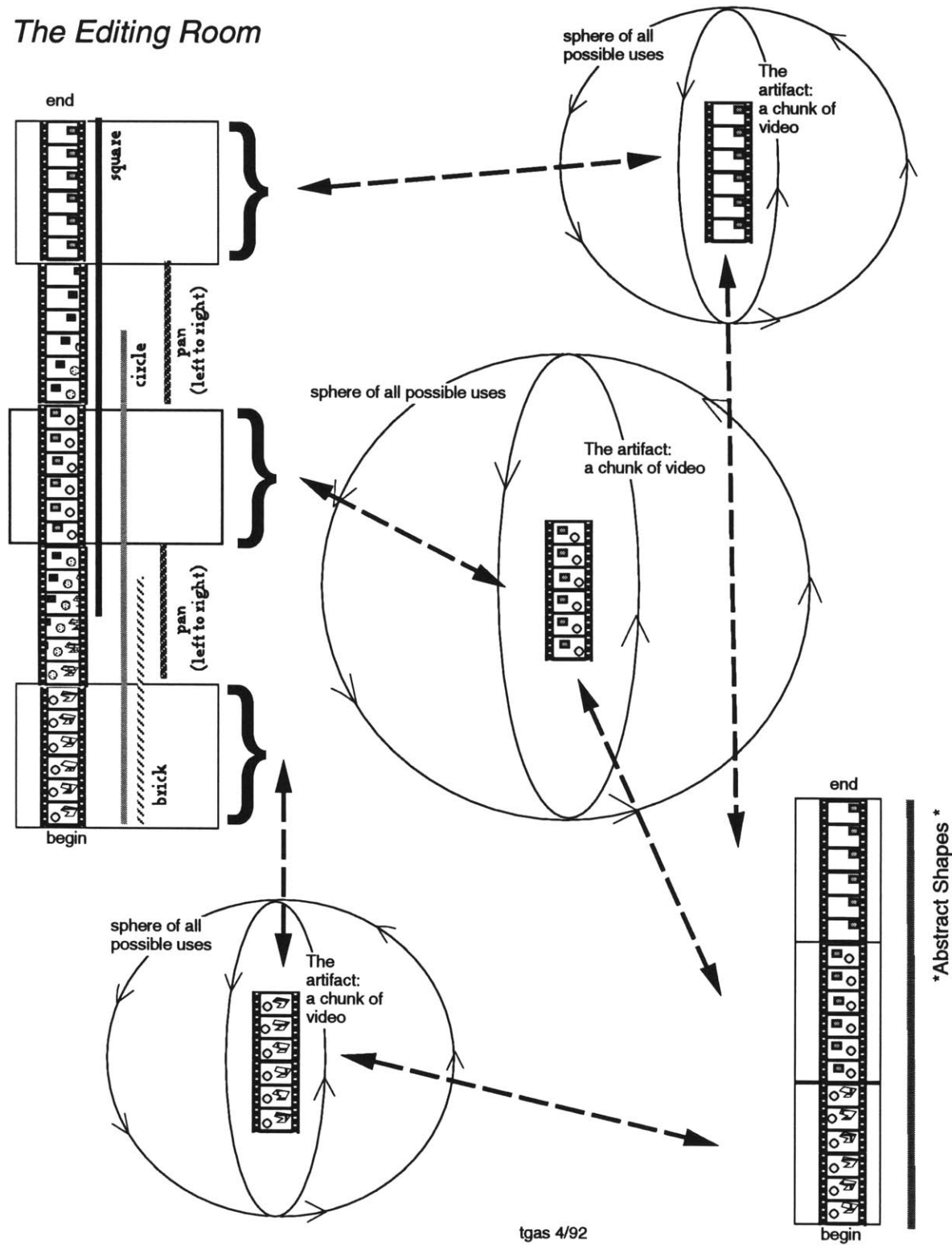
Once in the editing environment (figure 10), chunks are selected, de-contextualized and transformed into transitional objects. They are subject to the editor’s intentions, and moves as he reflects and acts to design the final edited work. In the motion picture design environment of the editing room, the recorded video tape has the status of being in a virtual “purgatory” of intentions. It is both the result of a set of intentions to shoot something and the raw material from which a new set of intentions arises. The motion pictures and sounds that have been recorded on the medium are the result of the film maker’s past shooting intentions that were grounded in real experience and now subject to future directed goals. The video must transcend the makers account even though it is a product of a personal experience.

---

<sup>1</sup>The way that these intentions are realized is dependent upon how the medium can be manipulated. For example, for analog video chunks are ordered in a linear sequence. In a digital design environment elements of an individual frame can be manipulated.

Figure 10: Chunks are of the video stream assembled into a new sequence called \*Abstract Shapes\*.

## The Editing Room

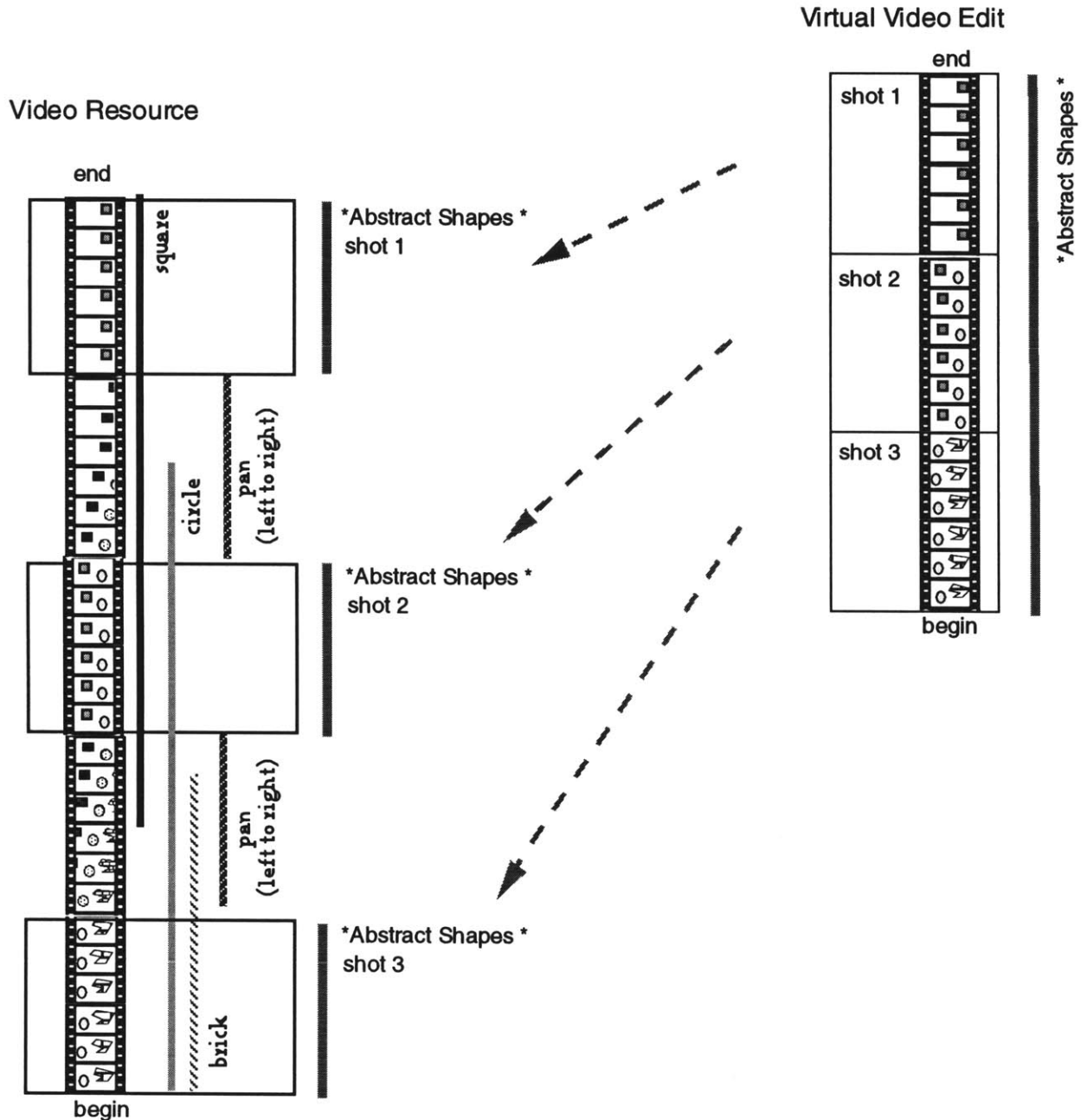


The sequence designer cuts out moves 2 and 4 (the pans) so that there is an abrupt jump from one shot to the next. In the sphere of possible uses the chunk of video becomes a reference entity that is pulled by both the past intentions of the camera person and future intentions of the editor. When placed into a new sequence - the new content "Abstract Shapes" emerges. The new sequence stands on its own; in some ways wipes out the original footage. The design environment is generative because each chunk of video can be the object of myriad intentions and as a result be placed in any context. Although the moves in this design space are constrained (here we are just talking about the reordering and chunking of sets of contiguously recorded frames) the motion picture artifacts are free to be placed in any context. As design knowledge is applied to the motion picture artifacts, new content and meanings emerge. A chunk of video has a trace back to its history of generation but also it is the site where additional intentions are directed. A recorded moving image is the nexus of intention vectors which point to its origin (the memories of how it was shot) and intention vectors which point to any future use. These future uses represent potential design moves that have not been realized yet.

#### *5.5.1 Stratification and Virtual Video Editing*

Creating an edit on random access system does not entail that a new video tape is created by copying the original source material. Here is where the break with conventional editing systems occurs. The edits can be virtually assembled at runtime - movies can be stored as a play list which consists as a set of pointers to the source material. These edit lists are a kin to other types of lexical annotations that are applied to the video stream during logging. When a new sequence is created the maker needs to name it so that he can later find it and play it back on the system. Figure 11 depicts how the sequence name can be used to reflexively annotate the video resource. Creating sequences becomes a passive form of data entry.

Figure 11: Sequence Annotations. Sequence annotations reflect a shift in significance of the video resource. In an edited sequence new context is provided for each chunk. This new context gives more information about what a chunk of video resource represents.



The virtual video edit is a play list that serves as another type of annotation for the video resource. In creating virtual edits, the maker is doing passive data entry. The new context for an edited sequence effects the meaning of the frames that compose it. In this way, the chunks of video resource become embedded in a web of past uses which are edit lists for virtual movies. As with the stratified de-

scriptions which were applied in the shooting environment, sequence annotation trace the video maker's intentions that are developed in the editing design space. Through the process of reflection in action new content emerges in the new chunking of video that the editor produces. These new chunks and their associated content are artifacts of all the past moves which occurred in the editing design space. The virtual edit lists tells other users how the meaning of those frames is transformed into new content in a virtual edit. They are also a trace of the makers intentions as they developed while interacting with the medium. This information can then be used as an index for searching. As shown in the figure, if a user were to search the lexical descriptions for brick then he would also notice that those same frames are embedded in the strata "Abstract Shapes -shot3". This strata line is a *dynamic link* to the virtual edited sequence "Abstract Shapes".

Stratification serves as a computer representation of the video stream on random access system computer system. The integrity of the context of where the footage was shot is maintained while additional contexts can be created during the process of sequence assembly. When a researcher annotates a segment of video, he is in effect creating a new stratum that is associated with a particular moving image in an edit. The old strata lines still exist in the database, but in editing a new meaning emerges in the re-juxtapositions of image units. This new meaning takes the form of an annotation. In a sense, the content of a series of frames is defined during logging. Yet the significance of those frames gets refined and built up through use. This information is valuable for individuals who want to use the same video resources over a network in order to communicate with each other. The stratification method provides a way to represent alternative "readings/significations/edits" of the same video resource to co-exist on the system.

## 6. *MAKING KNOWLEDGEABLE LEXICAL DESCRIPTIONS*

We have described a design environment where visual descriptions can be created and manipulated. The next problem is to find a way to organize lexical descriptions of the content of the video stream in a knowledgeable way. Usually in video production, logging is the most dreaded experience. When logging, makers only want to describe what they need to describe. They need to mark a particular chunk of the video stream so that they can find it when assembling a sequence. The types of description that are applied during logging are useful within the scope of a particular production. They reflect both the maker's memory of what was shot and the intention to find it later in the editing room. When taken out of this scope, a video log is often reduced to jibberish. Let's take a close look at this phenomenon as it relates to the Anthropologist's Video notebook and the stratification methodology.

To create an image is to move from external stimulus toward a medium. Similarly, to name an object (which can be an image) is to move from external stimulus toward language. From Marr we realize that a camera is a representational system which creates a description. That description is the image. We can extend this idea to the application of names. The naming of something is a representational system which creates a description. That description is a name. The representational system is directly related to the name it creates. For example, in the "boy with broom collecting trash" footage (Chapter 3 page 28) a botanist would be interested in the type of plant that was used as the broom. He would describe the chunk of footage (frames 92096 -93442) as an example of the use of the plant "Meste". While an urban anthropologist would describe the same chunk of footage as "child labor in Chiapas." Both the botanist's and the anthropologist's particular discipline functions as the system of representation for a lexical description. The naming of the object is only coherent within the scope of that particular discipline or representational system.

The way that a name is applied to an object (the way a system of representation generates descriptions) becomes a symbolic mapping from an object to a word. This mapping produces a single strand of meaning. As Marr explained earlier there is an "opportunity cost" when one chooses a particular representational scheme over another. Being committed to only one representation system can result in a great loss of understanding. In the example above, if we describe the video only in the botanist's terms we focus only on the broom and not the boy. Yet although many systems of representation are used to describe the same objects, people are still able to derive meaning from the words used to describe them. Many different types of representational systems create a complex network of understanding. The chunk of video depicts an example of "child labor in Chiapas" as well as the use of "Meste." Both these descriptions conflict with one another yet they are allowed to coexist within the video design environment.



The ambiguity of images lies in the fact that for any image there can be many words that correspond to it. How can we account for the fact that one object can have many names attached to it? The ambiguity of language confronts us every time we try to describe an object. How do we determine or decide which name to use? In the example above the descriptions given by the botanist and the urban anthropologist seem to be woven together for the same object. It seems that images are not so much ambiguous entities as their descriptions make them so.

Wittgenstein the *Philosophical Investigations* (1958) struggles with the same issues only from a different perspective. By focusing on language games and language understanding, Wittgenstein shows how the ambiguity of naming can be traced to a fundamental paradox of what it means to follow a rule for naming an object. Wittgenstein's task is to clarify the confusion caused by philosophers in their investigations of language. The issues that concern Wittgenstein are directly related to the problems of describing and interpreting video in a database.

### 6.1 *Moving Image Databases and the Ghost of St. Augustine.*

Wittgenstein opens *The Philosophical Investigations*, with an excerpt from *St. Augustine's Confessions*. Augustine writes "Thus, as I heard words repeatedly used in their proper places in various sentences, I gradually learnt to understand what objects they signified; and after I had trained my mouth to form these signs, I used them to express my own desires" (Wittgenstein, 1958 : § 1). For Wittgenstein this quote represents a particular picture of language where "individual words in language name objects -- sentences are combinations of such names. -- In this picture of language we find the following idea: Every word has a meaning. This meaning is correlated with the word. It is the object for which the word stands" [Ibid.]. Wittgenstein argues that St. Augustine's Picture Theory is not a good model for understanding language and the meaning of words. One can easily see in the example used above how this theory breaks down (the correlate of the description "Meste" is an image that also has the description "child labor" correlated with it).

Wittgenstein's critique of St. Augustine's Picture Theory has deep implications for the creation of the Anthropologist's Video Notebook. Upon considering the process of how words are applied to objects, the utility of Wittgenstein's critique will become evident. The process of correlating an object with a word is accomplished by "training." Wittgenstein calls this training "the ostensive teaching of a name."

An important part of training will consist in the teacher's pointing to the objects, directing the child's attention to them, and at the same time uttering a word; for instance, the word "slab" as he points to that shape. ... This ostensive teaching of words can be said to establish an association between the word and the thing. But what does this mean? Well, it may mean various

things; but one very likely thinks first of all that a picture of the object comes before the child's mind when it hears the word. ... Doubtless the ostensive teaching helped to bring this about; but only together with a particular training. With different training the same ostensive teaching of these words would have effected a quite different understanding (Wittgenstein, 1958: § 6).

Wittgenstein's account of this process is strikingly similar to the way that descriptions of images are entered into a computer database. Are the conventional approaches to describing objects in a video database system haunted by St. Augustine? The user of a video database enters a key word or some type of lexical description in order to find the corresponding image. Thanks to technological innovations, the images appear on a computer screen instead of in one's head as St. Augustine would have conceived of it. To ostensively define an object is to "pick out" a name from an object. *If we give the computer a lexical description we expect the computer to reconstitute it by presenting an image on the video screen. That video screen is the heir to the mind in St. Augustine's Picture Theory.*

To ostensively define something is not just to apply some rule that will correlate a word with the object. And here, according to Wittgenstein, lies the problem: "With a different training the same ostensive teaching of these words would have effected a quite different understanding" (Wittgenstein, 1958 : §6). Training plays a critical role in understanding. Wittgenstein further develops the point:

... when I want to assign a name to this group of nuts, he might understand it as a numeral. And he might equally well take the name of a person, of which I give an ostensive definition, as that of a colour, of a race, or even of a point of the compass. That is to say: an ostensive definition can be variously interpreted in *every* case (Wittgenstein, 1958 : § 28).

A different rule could correlate another word with the same object. The urban anthropologist might not know what "Meste" means. In this way, there could be an indeterminate number of objects that a name describes. Yet this leads us nowhere. Does the naming of an object always result in a paradox?

## 6.2 Language Games -- The Fibers and "Strata" of Understanding

What then do these objects have in common if they share the same name? A lexical description of a chunk of video in a database depends on descriptive practices or techniques. As discussed above, the process of understanding an ostensively taught name is related to the *use* of the name together with an *explanation* of the name's use. Wittgenstein calls these descriptive practices *language games*. In § 65, he further refines the notion of a language game. He asks, "What is common to all these activities, and what makes them into language or parts of language." And then replies,

Instead of producing something common to all that we call language, I am saying that these phenomenon have no one thing in common which makes us use the same word for all, but that they are related to one another in many different ways. It is because of this relationship or these relationships, that we call them all “language” (Wittgenstein, 1958: § 65).

This brings us back to the previously mentioned assertion that it is less the image that is ambiguous but more the language that surrounds that image. Wittgenstein cautions us that if someone was to *look* and *see* what is common they “will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. ... (We) can see how similarities crop up and disappear. And the result of this examination is: we see a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail. ... I can think of no better expression to characterize these similarities than family resemblances” (Wittgenstein, 1958 : 66-67). These similarities which overlap each other form a complex weave of ambiguity. This weave of family resemblances is what Stratification captures and allows us to develop in the video design environment.

Understanding a language is not dependent upon one canonical system of rules that link objects to words. Attempts to describe video using a consensus representation that relies upon what can be agreed upon among many individuals are a kind of demographic naming of objects that does not take into account the infinite possibilities of naming or how video might function in a design space.

Rather, Wittgenstein’s language game, with its rules and techniques for correlating words to objects, produces one *fiber* of meaning for each correlation. The understanding of a word is captured however, by many correlations of meaning i.e. by a *thread* or *strata* of many fibers.

...in spinning a thread we twist fibre on fibre. And the strength of the thread does not reside in the fact that some one fibre runs through its whole length, but in the overlapping of many fibers (Wittgenstein, 1958 : §67).

The strata of meaning has two sources — a language game for many words generates many related fibers, and in addition, many language games can contribute their own fibers for the same chunk of video. In this way name is piled upon name, meaning is piled upon meaning producing a memory of *all* the previous incarnations of the image. Through analyzing these layers we can begin to understand the significance of an image in it’s constantly shifting web of descriptions. Understanding the image, however is not just analyzing the single frame within its over lapping fibers, but rather we must take into account that something that “runs through the whole thread namely the continuous overlapping of these fibers” (Wittgenstein, 1958 : §67). We must understand an image in relationship to its entire web of concurrent meanings.

Every time we name something we choose a language game. The ostensive teaching of a name is a type of language game that involves pointing and uttering a word. When one changes the rules then it also follows that the nature of the language game correspondingly changes. Thus, ostensive definitions that describe objects in the external world are consistent only within the language game in which they were generated. (The botanist with his “Meste,” the urban anthropologist with his “child labor in Chiapas.”) The precision of generating a name from an object is misleading because the name generated is relative to a particular language game or discipline.

Wittgenstein describes this tension between the exactitude of a naming procedure and the arbitrariness of a name:

Naming appears as a queer connection of a word with an object. And you really get such queer connection when the philosopher tries to bring out the relation between a name and a thing by staring at an object in front of him and repeating a name or even the word “this” innumerable times (Wittgenstein, 1958 : §38).

The understanding gained from a name considered in isolation is both relative and fragile. Only a single fiber of meaning is created during the process of naming. Only upon appreciating that these individual fibers are part of an ever expanding thread of understanding do we get out of the loop of staring at an object and repeating the name ad infinitum.

Wittgenstein’s aim is to tease out the problems that are manifested in the *process* of naming an object. To name a chunk of video is to produce a fiber of meaning. Understanding the significance of a chunk of video does not depend on any one naming procedure but instead arises out of a strata of many different fibers of meaning. To understand a language (in this case a visual language), is to account for ambiguous and contested meanings. To account for the entire web of meaning.

As the strata lines overlap we don’t just have the logical conjunction of all the related descriptions we have something more valuable -- a computational representation for ambiguity. We can visually examine the way that different language games (lexical representation systems) pull and contest the visual descriptions of an observed event.

The stratification system is a way to represent dynamic data. The video stream is a description that has been produced by the representation system of the video camera. The lexical descriptions of the content of the video stream are also the result of a particular descriptive strategy/language game representation system.

### *6.3 Implications for the Automatic Logging of a Video Stream*

Many different types of descriptions can be integrated in to the strata for a video stream. First, there are the sounds and images themselves which have been inscribed onto the medium by the camera as a representation system. Next, there are the various discipline dependent language games which generate lexical descriptions. In addition there can be computer programs that can each contribute descriptive fibers.

In 1985 a computer program which attempted to segment video material was built by Sasnett(1986). The system called "Scene Detector" flagged changes in image content by examining luminance and chrominance levels for nine points distributed on the video screen. Via these points the program would algorithmically determine if there was a scene change. A more recent version of this method is implemented in the Video Streamer (Elliot, 1992). Teodosio (1992) developed algorithms to detect and later manipulate pans, zooms of a digital video signal to create Salient Stills. The automatic detection of scene changes and transitions could allow the maker to concentrate on other types of lexical descriptions which are focused on the particular production. The attributes that are detected by these logging utilities could be easily correlated with other types of descriptions via a time code number.

Pincever (1991) describes how the combination of the determination the spectral signature of a sound with the lexical description of that sound can be used to create a template. In turn these templates could be used for automatic logging:

Such templates can be built for a number of different sounds which are usually encountered in most homes movies: cars, airplanes, etc. Thus, it will be possible for the system to not only recognize shot boundaries, but also to catalog some of the content of the shot as well. This will allow the system to create a log of the raw material. Also, this can lead to the implementation of a "search" function, that would be able to search through the material to find specific sounds. Suppose a user wants to find a shot in which there's a car. She would then call the search function, with "car" as an argument. The search function could then find the template of a car, previously recorded and analyzed, and run it through the material until a match is found. Then, the shot containing the car would be played back to the user. This would provide an initial shot description (Pincever, 1991: 36).

Turk (1991) developed a computer system that can locate a subject's head and then recognize the individual by comparing face characteristics of known people. Furthermore, he was able to detect the direction that some one is gazing. These computationally derived gaze vectors can be useful in editing. In a movie when a character looks to the right and then there is a cut to an object a spatial relationship is articulated between the character and the object.

Such applications will eliminate the drudgery of describing video. In the future, the maker can concentrate more creative aspects of video production rather than on bookkeeping tasks. Although the possibility of automatically creating a database of the content of video stream is seductive, it should be remembered that the content of the video stream dynamically changes over time. That fact that chunks of video are arranged into a sequence reveals how those frames are interrelated. Stratification is a representation that can both account for how visual and lexical descriptions change within the video stream and how they change over time as they are used.

## 7. MAKING THE ANTHROPOLOGIST'S VIDEO NOTEBOOK

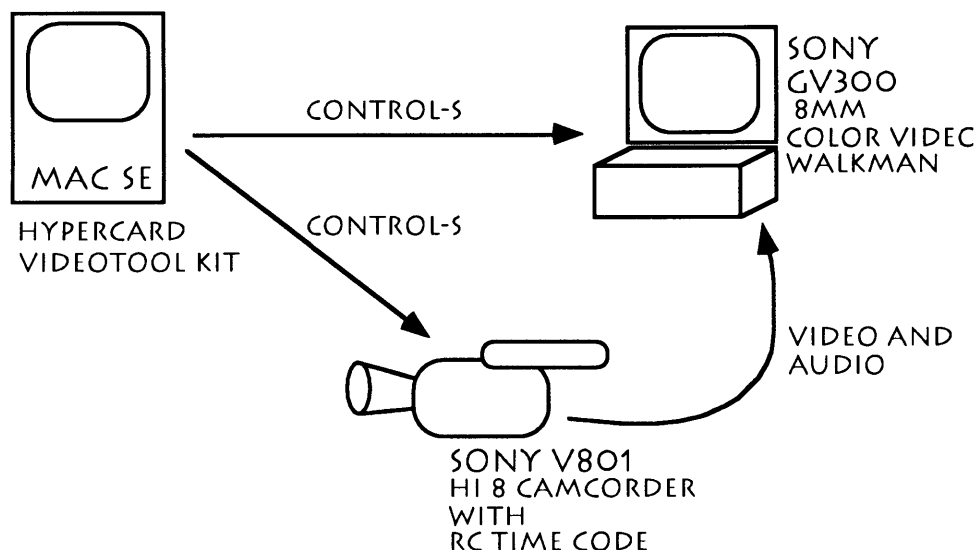
The Anthropologist's Video Notebook was developed in two stages which reflect the needs of two distinct research environments. In Mexico, a field notebook was created to aid in the acquisition of textual descriptions of recorded images. The field notebook was developed as the field research progressed. The design was iterative, when modifications were required they were made. Much of the design knowledge gained in this practical research setting was integrated into a video database research environment called the Stratification system. The stratification system was implemented on networked DECstation 5000 UNIX workstation. Here, emphasis was placed on the reusability of video and on the description of the content of video so that different researchers, each perhaps having different interpretations of the same material, can see what they want. The field notebook was for data gathering and the Stratification system on the UNIX workstation was for analysis and sharing of results.

### 7.1 *The Field Notebook: System Configuration*

Anthropologist's field notebook consisted of Apple Computer's Macintosh SE computer, Sony's V801 camera with RC time code and the Sony GV300 color 8mm video watchman. The V801 and the GV300 together with video tape cassettes, batteries and lenses fit into a medium sized book bag. I could comfortably carry all the video equipment, sleeping bag and clothing with ease. The compactness of the set up allowed me freedom of movement when visiting remote villages. I used the V801 to record video and the GV300 to play back what I just recorded to small audiences. These impromptu screenings gave me an opportunity to concretely demonstrate what I was shooting and gave the community an opportunity to evaluate my work. The screenings enabled me to demonstrate what I was recording to the healers and helped communicate my goals and intentions of the project to them. By showing the footage in the original setting I was able to incorporate suggestions and make shooting a reflective process.

Back at the PROCOMITH research center in San Cristobol de Las Casas, an Apple Macintosh SE computer was attached to the V801 and GV300 (see figure 12). Abbate Video Consultants' *Video Tool kit* (1992) provided the interface between the Macintosh SE and the video equipment. Video Tool kit consists of a special cable which connects the V801's control-S jack and the GV300 control-S jack to the Macintosh modem port. In addition to the cable, Video Tool kit includes a Hypercard application for video logging and assembly called *Cue Track* and a set of commands that enable the Hypercard stack to control video devices via the cable.

Figure 12: Field Notebook Configuration



With this set up I could control the V801 to cue up in and out points and also view these clips on the GV300. Since Hi8 camera was equipped with RC timecode<sup>1</sup> frame accurate annotations were attainable. For logging the GV300 was primarily used as a monitor because it was not equipped with RC timecode. Frame accuracy was a necessity -- for this reason the V801 was used exclusively.

This setup also allowed for bare bones assembly video editing. For editing the V801 was used as the video source (due to its frame accurate RC timecode) and the GV300 was used as the record deck. The set up allowed for the production of rough edits but fell short of high quality production needs because the GV300 can only record in 8mm format not Hi8 and is not frame accurate. The correct frames are cued and played by the V801 but the GV300 can only record in record - pause mode. A GV300 equipped with RC timecode and Hi8 recording ability would provide the accuracy and quality needed for in-the-field video production.

---

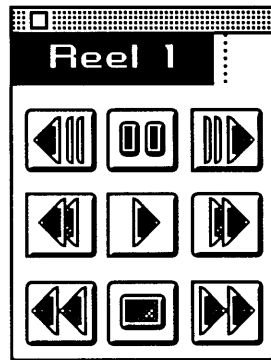
<sup>1</sup> RC Timecode is Sony's "special" consumer timecode. "RC" stands for Re-write able Consumer time code. Unfortunately, RC time code is not compatible with the time code available on professional video editing machines. RC timecode is not even compatible with the timecode used on their professional Hi8 editing machines (EVO9800 and the EVO9850)! Though useful for frame accurate logging in the field, RC timecode requires re-stripping the video tapes with professionals Hi8 timecode and calculating the offset between the two timecodes.



## 7.2 Logging Video with CueTrack

CueTrack's integration of video device control with Hypercard facilitated the logging process. The interface for controlling the video devices is a familiar video palette (Figure 13).

Figure 13: OnTrack's video control palette.



The V801 was controlled on screen by using the mouse to press the buttons on the video palette. CueTrack represents each videotape as a Hypercard stack consisting of cards. These stacks contain cards that represent shots or clips.










The original cue track application supported interactive entry of in and out points, and three types of descriptions: a clip title and two text fields called "Audio" and "Video." To create an annotation the video tape would be cued to a desired section using the video control palette and then grab the time code number of the current frame from V801. The time code is entered into either the in-frame or out-frame field and then text is added to the fields on each card.

The limitations of CueTrack became apparent as the number of video tapes increased. The main issue that I faced was that my subject matter was not confined to a single tape. Events and themes were often spread out over many video tapes. Given CueTrack's representation of an individual video tape as a stack, I could consistently find information within the scope of one stack. Since I was shooting many hours of tape, it soon became apparent that I needed a set of annotation tools which could help me describe and keep track of shots over the entire set of video tapes. To meet these needs, I modified the CueTrack application and created a set of additional Hypercard stacks.

### 7.2.1 CueTrack Stack Modifications

A sample CueTrack card that has been modified for use as the field notebook is presented in figure 14. The extensions to Cue track are discussed below.

Figure 14: Video Notebook Extensions to CueTrack.

tape14 (Mexico)																								
1	1.0.6		6/14/92 2:31:49 PM																					
2	Clip: using tobacco		  																					
3	<b>Description:</b> Se usa cuando toman trago o se rezan en el cruz. Es un tratameinto para el dolor de estomago. Cuando esta masticando el tobacco se da mas fueza para lograr la curacion. It is used when drinking or praying. It's a treatment for stomach ache. Chewing tobacco increases healing		<b>Audio:</b> Alonzo: Vamos a recibir el tobacco.  Alonzo: We are going to take tobacco.																					
4	key words	[ people : Sebastian Alonzo Brent ]		 																				
5	seqs:	[Healing Ceremony]		 																				
<table border="1"> <thead> <tr> <th>Times:</th> <th>Real Time:</th> <th>LTC:</th> <th>Frames:</th> <th>Counter:</th> </tr> </thead> <tbody> <tr> <td>In:</td> <td>00:32:04:19</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Out:</td> <td>00:32:20:09</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Dur:</td> <td>00:00:15:20</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>					Times:	Real Time:	LTC:	Frames:	Counter:	In:	00:32:04:19				Out:	00:32:20:09				Dur:	00:00:15:20			
Times:	Real Time:	LTC:	Frames:	Counter:																				
In:	00:32:04:19																							
Out:	00:32:20:09																							
Dur:	00:00:15:20																							
<b>Utilities</b>				<input type="checkbox"/> <b>Critical</b>																				
6	7	8																						

1 Date and Time Stamp: Each time a new card was created the date and time stamp is automatically entered into this field. The need for this field arose as I started to go over the tapes a second or third time. During these additional viewing sessions I would usually see something new and want to

create a card for the new observation. The time and date stamp allowed me to distinguish between old and recent observations. My ideas changed and interpretations changed over time the date and time stamp helped me track these changes.

2 Clip field: I did not make any changes to this field. In most cases the clip field and an associated in-point are sufficient for briefly annotating a chunk of video. The first pass through logging primarily consisted of defining a clip with an in-point and then adding text to the clip field. These bare bone annotations served as an mnemonic that were later incorporated in thick descriptions. An in-point and the Clip title is defined as a “content marker”.

3 Description and Audio Fields: Cue tracks original “Video” field was replaced by a “Description” field. The Description field was used to create “scratch notes” about the content of video. The description field included not only descriptions of the action but also explanations and personal insights. The Audio field from the original Cue Track stack was retained. It was used for transcriptions and translations of dialogue.

4 Keyword Field: The keyword field is where more generalized keywords for each clip are entered. The topic of keywords will be covered in the discussion of the keyword stack.

5 Sequences: The sequence field is a place associating a particular clip with a story thread. These story threads are called sequences. Sequence annotations will be discussed in depth in reference to the Sequence stack.

6 Utilities Button: The utilities button pops up the utilities stack in a new window. The utilities stack contains general “house keeping” functions that were required across all the stacks/tapes. It contained scripts which removed extra spaces and new lines from the text fields for each card was used to export data out of hypercard into delimited ASCII text files. With the time and date stamp, I could download all the annotations made after a particular date.

7 Keyword Button: Pops up the keyword stack in a new window. Since the screen real estate of the Macintosh SE was limited, the Keyword stack would only appear when needed.

8 Sequence Button: Pops up the sequence stack in a new window.

### *7.3 The Keyword Stack*

Keywords are used to keep track of people and objects. They are a more generalized type of description that is used across many different video tapes. They are organized in a two level hierarchy of keyword class and keyword. Each keyword class is represented as a single card (figure 15). Each card

contains a field for all the key words in that class and a field to hold all the keywords that are currently in use. In the example below, the key word class is people and the list of names are the keywords that are associated with that class. The text field at the bottom of the card indicates the keywords that were used for the previous clip.

Figure 15: Keyword Stack for People Keyword Class

**keywords**

**people** \* **Carmelino**

Sebastian  
Alonzo  
Martha  
Luisa  
John  
Thomas  
Brent  
EAB  
Victor  
Tere  
Alonzo  
Sebastian  
Manuel  
Guadalupe  
Collaborators  
Carmelino  
Feliciano

↔ → ↶

**Add to List** **Send to Clip**

**List of current keywords:**

[ people : Sebastian Alonzo  
Brent EAB ]

### 7.3.1 Using the Keyword Stack

In observational video the place as well as the characters present in the environment remain constant for prolonged periods of time. The keyword stack facilitated the consistent entry of keywords and also saved time and work that were required to redundantly describe the keywords that were in effect for each content marker. Here it should be pointed out that during this stage of the research I used keywords as a supplementary form of annotation. I first created a content marker that consisted of a clip name and an in-point and then would quickly add the list of key words that were in effect for that content marker. Since a set of keywords are usually in effect during a whole series of content markers, the list of previously used keywords was saved in the current keyword field. When a descriptive state

changed --say, someone entered the room -- the current keyword list could be correspondingly modified by just adding another name. The keyword utility was implemented toward the end of my time in the field. For this reason, it was not used extensively. The utility of keyword classes really comes to the fore in the workstation environment.

#### 7.4 The Sequence Stack

The sequence stack provides a place to develop story intentions (Figure 16). The sequence stack guides my intentions during the shooting process. Each card in the sequence stack represents a possible story element or thread. It serves two purposes. First, it is a way to keep track and organize ideas or key themes about the video footage. Second, it is a simple database of how different chunks of video are related to each other.

Figure 16: The Sequence Stack-- Thematic Category: Establishing San Cristobol.

<b>Sequences</b>			
<div style="display: flex; justify-content: space-between; align-items: flex-start;"> <div style="width: 20%;"> <p><b>Title</b></p> </div> <div style="width: 80%; border: 1px solid black; padding: 2px;"> Establishing San Cristobol </div> </div>			
<p><b>Description of Sequence :</b></p> <div style="border: 1px solid black; padding: 5px; min-height: 60px;"> The Market  street cleaners early in the am  church colors on the outside  ... </div> <div style="float: right; text-align: center;"> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↑</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">□</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; background: repeating-linear-gradient(45deg, transparent, transparent 2px, black 2px, black 4px); cursor: pointer;"> </div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↓</div> </div>			
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p><b>Need to shoot</b></p> <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"> ↔ ↔ ↺ </div> <div style="border: 1px solid black; padding: 2px; flex-grow: 1;"> Trash Collectors  Festival of San Jose  Policemen  Taxi Cabs </div> <div style="margin-left: 5px; text-align: center;"> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↑</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">□</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↓</div> </div> </div> </div> <div style="width: 45%;"> <p><b>Continuity/Conditions</b></p> <div style="border: 1px solid black; padding: 2px; min-height: 40px;"> Rooster Crows </div> <div style="float: right; text-align: center;"> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↑</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">□</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↓</div> </div> </div> </div>			
<div style="display: flex; align-items: flex-start;"> <div style="width: 15%; text-align: center; margin-right: 10px;"> <div style="border: 2px solid black; padding: 5px; width: 40px; height: 30px; margin: 0 auto; position: relative;"> <div style="position: absolute; top: 0; left: 0; right: 0; border-top: 2px solid black;"></div> <div style="position: absolute; bottom: 0; left: 0; right: 0; border-bottom: 2px solid black;"></div> <div style="position: absolute; left: 0; top: 50%; transform: translateY(-50%); border-left: 2px solid black; width: 10px;"></div> <div style="position: absolute; right: 0; top: 50%; transform: translateY(-50%); border-right: 2px solid black; width: 10px;"></div> </div> <p style="margin-top: 5px;">grab shot</p> </div> <div style="width: 85%;"> <p><b>Shot Location :</b></p> <div style="border: 1px solid black; padding: 5px; min-height: 100px;"> [tape01,on bus to SC, 00:23:12:18, ]  [tape03,view from roof top, 00:57:35:27, ]  [tape03,woman doing laundry, 00:57:49:19, ]  [tape04,wandering at the market, 00:49:58:23, ]  [tape13,vamos a rezar aqui, 00:10:08:27, ]  [tape23,Catarina with son, 01:25:46:01, ]  [tape08,child on back, 00:32:25:09, ] </div> <div style="float: right; text-align: center;"> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↑</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">□</div> <div style="border: 1px solid black; width: 20px; height: 20px; margin: 2px; cursor: pointer;">↓</div> </div> </div> </div>			
<div style="display: flex; justify-content: space-between; align-items: flex-end;"> <div>1.0</div> <div style="font-size: small;">Note : to print report do "print report" in File menu.</div> </div>			

The sequence stack provides a quick way to check and review the different story elements. For example, when I was shooting Carmelino with the intent of showing how he commutes from the San Cristobol to Chamula I found myself in the market. While in the market, I knew that I needed to get a

shot of livestock. I took advantage of this opportunity and was able to shoot chickens and pigs in the market although my intent for being in the market was to shoot Carmelino's daily commute. One intention gets me into a particular location but as the action unfolds I need to be cognizant of opportunities to shoot something that satisfies another story intention. By reviewing the sequence stack, I could keep abreast of the many different types of story threads and make the most out of shooting - it was a way to keep these story ideas fresh in my mind. I could always check to make sure I was getting the shots that I needed in order to tell a story that I might want to create in the editing room.

#### *7.4.1 Using the Sequence Stack*

The sequence stack has various text fields: a title field, a field for describing the content of the sequence, a field for keeping track of what still needs to be shot and a field for continuity/conditions which might pose problems while editing. The "shot location" field displays all the tape locations where a particular sequence has been used.

Like the Keyword stack, the Sequence stack is accessible via the CueTrack stack. When the TV icon is clicked, the sequence stack appears in its own window. To find a desired sequence, I flip through the various cards of the sequence stack. When the grab shot button is activated, two things happen. First, the sequence stack sends the title of the sequence to the CueTrack card I am currently working with and places that data into the sequence field (refer back to figure 14). Then, the sequence stack retrieves the shot location (tape name, title, in-point, out-point) from the current CueTrack card and places this information into the shot location field of the sequence stack. Items in the shot location field can be double clicked and the corresponding CueTrack stack for that tape appears. At this point, the actual video tape can be loaded and cued to the associated in-point.

#### *7.5 Making Translations*

The most important benefit of having a database of descriptions integrated with video recorders was the ability to translate and transcribe dialogue while in the field with a native speaker. Once key chunks of dialogue were defined, CueTrack facilitated the play back of a sequence over and over again to get an exact translation from a Mayan Collaborator. Often when using recorded audio tape the process of going forward and rewinding is not only disorienting for the researcher but also for the indigenous translators alike. The addition of video to recorded audio augmented the recorded dialogs. While translating with a Mayan collaborator, the video provided a visual context to the recorded utterances.

By translating while in the field I could ask more questions concerning a specific point that I did not understand. If I had waited until I was back in the United States, this information would be impossible to obtain. Figure 17 illustrates how new insights about the content of the video were gained

during translation. I logged this chunk of video in two passes, the first was for bare bones content information and the second was for translation with a Mayan collaborator. While translating, with a Mayan collaborator (the segments created on 8 August 1991) I was able to augment my initial scratch notes (the content markers created on 15 July 1991) with thicker descriptions. This chunk of video was shot right after a healing ritual. During the ritual the healer Sebastian (Seb) was required to pray for one hour while on his knees. When done he takes a sip of Posh but before he does he must say “Cheers” to every one else in the room. He then asks his daughter for a chair and then he passes the cup around to room so that other people drink. In two logging passes my understanding of the content has grown substantially. The last three content markers are contained in the segment beginning at 00:02:13:00. Although the video is described at two different times and for two different purposes, the content markers provide valuable contextual information about the dialogue.

Figure 17: Example: Translation with further annotation .

Tape	In Frame	Out Frame	Shot Name	Description	Audio	Date
tape15	00:01:33:23		drinking posh			7/15/91
tape15	00:01:54:00	00:02:12:07	trans 1.55	Posh is home brewed sugar cane alcohol. Which is used in healing ceremonies and just about any other time as well.	Seb: Saludo; Al: Saludo, senora MV: Saludo; Seb: se dolio mi rodilla, traeme una silla hijita; Hija: Bueno  <i>English:</i>  <i>Seb: Cheers</i> <i>Al: Cheers</i> <i>Seb: My knees hurt. Go get a chair for me daughter</i> <i>Daughter: OK</i>	8/8/91
tape15	00:01:54:17		Alonzo drinking			7/15/91
tape15	00:02:13:00	00:02:27:01	trans 2.14.	There are different ways to say cheers when drinking Posh  X'ochon jtat (o'shon tat)= salud, hombre <i>Cheers to Men</i>  X'ochon jme'nin (o'shon men in)= salud mujer <i>Cheers to Women</i>  X'chon bankilaetik (o'shon ban ki la le tik ) = saludos a todos los hombres <i>Cheers to Everyone</i>	MV: salud Todos: Saludo  <i>MV: Cheers Everyone</i> <i>Cheers</i>	8/8/91
tape15	00:02:16:01		Sebastian sitting down			7/15/91
tape15	00:02:26:19		girl in red drinks			7/15/91
tape15	00:02:53:13		other guy drinks			7/15/91

## 7.6 *Getting the Big Picture*

The clumsiness of Hypercard became apparent when it came time to search, sort and print out formatted reports of the annotations for all 27 hours of video tape. To accomplish these tasks with the CueTrack, 27 different Hypercard stacks would have to be opened and manipulated. To get around this problem, I dumped all the data out of the CueTrack stacks and uploaded them into FilemakerPro, a flat file database. Shots were easily found with Filemaker Pro. When the desired shot was located, I loaded the appropriate tape into CueTrack for automatic cueing.

### 7.6.1 *Of Time Codes, Video Editing and Switching Computer Platforms*

The purpose for the initial list of annotations was to direct me to the appropriate spot in the raw footage. Once back at the media lab, this database of annotations was used to find shots for a 20 minute video tape called "Mayan Medicine in Highland Chiapas Mexico." This video tape was created by first making a window dub of raw tapes<sup>1</sup> and then using this window dub version to produce a fine cut. In creating this fine cut I was creating another type of database of shot locations which would be used to assemble the final version of "Mayan Medicine." The final master edit was in turn put onto a video laser disc so that virtual edits could be created on the UNIX based Stratification system.

The purpose of the final edit list is to direct the assembly of the final video tape (Sasnett, 1986). Now, these lists can be thought of as two types of databases - one reflects my interpretations of the content of the footage while logging and the other reflects my moves that lead to the construction of new edited content in "Mayan Medicine." Usually the existence of two distinct databases results in the breakdown in the transmission of the original annotations. The list of initial annotations is essential for locating the shots that were delimited and assembled into the final version of "Mayan Medicine." These annotations provide a valuable key into to the raw footage. Furthermore, the fact that some of this raw footage was incorporated into video is also important. In "Mayan Medicine", relationships between various chunks of raw footage of the 27 hours of raw footage are made explicit. For the most part, this information is not reincorporated back into the database of description of the raw footage where it could be used by other video makers. Sasnett explains that this problem stands in the way of the development of reconfigurable video:

---

<sup>1</sup> A window dub is a copy of the original video tapes with the timecode numbers visible.



If these in/out numbers could be married to a larger data set which described the segments in human terms, especially their subject matter or content in some sort of systematic fashion ( as in a database), we would effectively have a computerized index of the video tape material. Since an edit list must be constructed to produce almost every master tape, it seems senseless not to make an enhanced version of this data available to end users. Otherwise, someone will have to recreate the edit list (by logging the materials) in order to reconfigure the video, and none of the information created by the makers passes through to the users (Sasnett, 32).

Since I had data gathering tools available with me in Mexico, it was possible to experiment with ways to pass this information on to users of UNIX based Stratification system. To maintain my initial descriptions of the events I was able to merge the edit decision list with the annotation database. The frame number of this merged database were then offset to coincide with the frame numbers on the laserdisc. In this way, I could maintain the descriptions that I had generated in Mexico with the field logging set up and use them in the DECstation 5000 virtual video editing environment. The resulting edit list with descriptions saved much work when creating a database for the video material on the laserdisc. I did not have to re-log the laserdisc to locate shot boundaries. Below is a segment of the edit list that was used to create the "Mayan Medicine" video disk (figure 18). In this example the information from the Macintosh based logs has been merged with the edit decision list.

Figure 18 Edit Decision list with database merge.

Edit #	VIDEO/ AUDIO	Source Tape	InFRM	OutFRM	laserdisc in~out
70!	V	11	14579	15653	[21651~22725]
		11	14903	antes no sabia	
		11	15373	pero ya pues	
71!	V12	16	16084	16466	[22725~23107]
72!	A12	11	14384	15958	[21456~23030]
		11	14459	favor de dios	
		11	14903	antes no sabia	
		11	15373	pero ya pues	
		11	15673	my eyes were closed	
73!	V12	12	3235	3521	[23107~23393]
		12	3381	carmelino looking at plant	
74!	V12	12	6083	6607	[23393~23917]
		12	6309	is this the plant	

The old descriptions don't provide enough information with which to discern the new content of this edited sequence. These initial descriptions which reflect my understanding of the footage while

in Mexico but don't reflect the new meanings that arise when the raw footage is placed in an edited context. The last column shows the correspondence to the frames on the laser disc.

In this example we have Carmelino walking down a trail into the sunset in Edit 71. This is a video only edit. Edit 71 has no descriptions attached to it because it was a cut which fell in-between the bounds of two content markers. Edit 72 an audio edit of Carmelino talking about his work as a Mayan collaborator for the PROCOMITH Project. Although the audio and the video were shot on two different days, in the editing room I developed the intention to place them together in a sequence. The content of the edited sequence is something new that I developed long after leaving Mexico. The logs from Mexico were instrumental in locating these two shots because they helped me remember the context of shooting. The usefulness of these descriptions when applied to the new edited context is diminished because they are out of scope. The edited sequence requires a new type of annotation that is consistent with intent of the maker and the context of the production of a final movie. The edited context provides us with more information and knowledge about that chunk of video. The question then is: How can a video maker's intentions be made useful for other people who want to reuse the video material. In the UNIX workstation environment, I experimented how the new content of assembled video sequences can be reconfigured by other users. On such systems, the linear context of a sequence of video is virtual, and as such, so are the descriptions of content.

## 7.7 *The Stratification System - A Design Space for VIDEO*

Conceivably, someone else can re-assemble the sequences of the “Mayan Medicine” video into another type of movie. Moreover, these “re-edited-versions” can later be used by someone else to make another video production. The process of editing using Stratification becomes the process of creating context annotations, and storing them along with the initial descriptions made during recording. The Stratification System is a discursive space where thick descriptions can be built for the content of video. Using the stratification system, the anthropologist can create thick descriptions from his video “scratch notes” that were created in the field notebook. The Stratification system enables researchers to semantically manipulate video *resources* and allows for alternative “readings / significations / edits” of the same material to co-exist.

Video resource can exist in many different contexts and in many personalized movie scripts. In a sense, the content of a series of frames is defined during logging. Yet the significance of those frames gets refined and built up through use. This information is valuable for individuals who want to use the same video resources over a network to communicate with each other. As we have seen Stratification is a descriptive methodology which generates rich multi-layered descriptions that can be used to trace the design moves and intentions behind a video production. Knowledge about video resource is built up through use.

The UNIX Stratification system allows for multiple users to annotate video using free text descriptions and more structured types of descriptions called keyword classes. There is an interactive graphical display of these keyword classes over time called the Stratagraph. Additionally, the system also allows for the assembly of sequences. The Stratification system provide technological support of video ethnography.

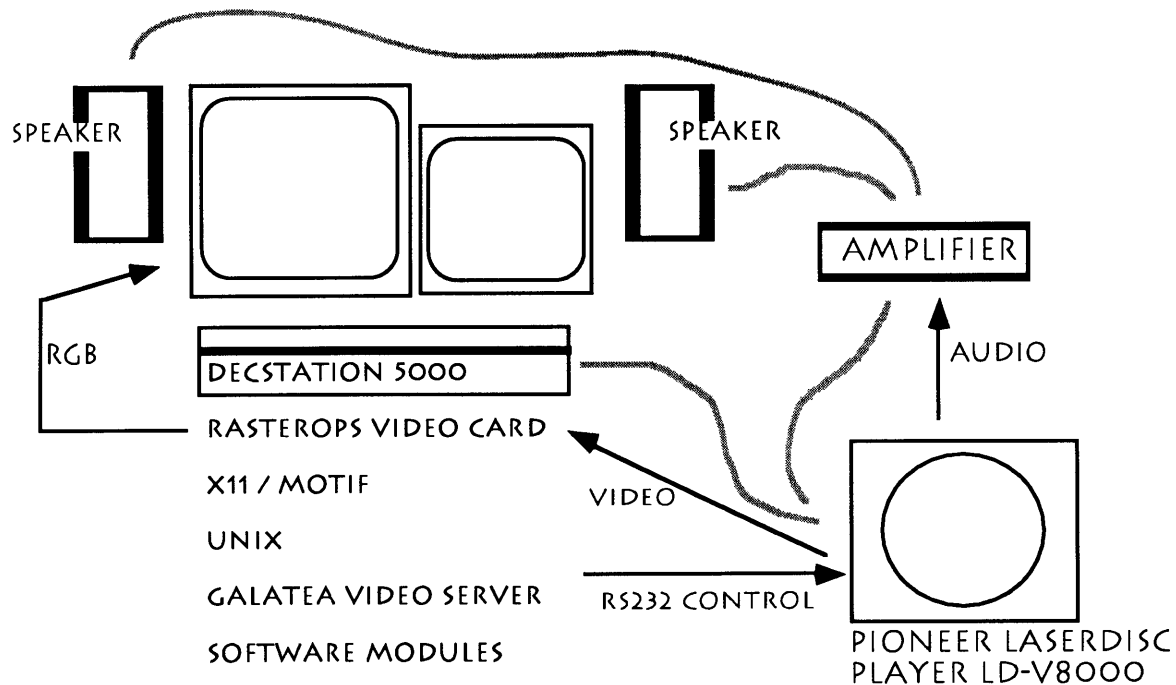
### 7.7.1 *System Configuration*

The workstation configuration for the Stratification System consists of the a DECstation 5000 computer equipped with a RasterOPs video/graphics board that allows live video to be digitized in real time and displayed on the workstations high resolution color monitor. Digital computer’s XMedia tool kit with its video extensions to the X Window System enables digitized video from the RasterOPs board to be displayed in an Xwindow.

The video source for the system is a Pioneer 8000 (a high speed random access laserdisc player) which is controlled via the serial port by a video device control server called Galatea (Applebaum, 1990).

The different modules of the stratification system are client applications that issue commands to Galatea. Galatea translates these abstract commands into Pioneer's device control protocols which are sent to the laserdisc player. The final piece of hardware is an amplifier and a set of speakers which is hooked up directly to the laser disc (figure 19).

Figure 19: Workstation Configuration.



## 7.8 Data Representations

The use of keyword classes and a special format for saving descriptions of video called Strata Data Format (SDF) are key features of the Stratification system. The implementation of keyword classes and SDF is designed to complement the file management and text processing utilities currently available in the UNIX operating system.


### 7.8.1 Strata Data Format

Each descriptive stratum consists of the source name, begin frame, end frame, free text description field, and keyword classes field. These descriptions are saved in delimited ASCII text files and stored in UNIX directories. SDF files are named in regard to a particular project and owned by an individual

or group like any other UNIX file. SDF files can be combined and analyzed for associative browsing of content across projects.

The first line of a SDF file is a special header which tells the location of all keyword classes that are used in the file. (Keyword classes will be discussed later). The strata data format is ASCII text file with fields delimited by “|” characters with one record per line. The fields are “Source | In- frame | Out- frame | Content- frame | Speed | Free text description | Class | Keyword | Class |Keyword”. The “source” is the name of the video source. “In frame” and “out frame” are self explanatory. The “content frame” is the most representative frame which is selected by the user. If no content frame is given then the median frame is placed in this field. The “free text description” that is usually a narrowly focused description that serves as a mnemonic. After the “free text description” there can be any number of “Class” - “Keyword” pairs. The first five lines<sup>1</sup> of the Strata Data Format File for the “Mayan Medicine” video disk are shown in figure 20:

Figure 20: Header and Strata for Mayan Medicine Videodisc.

<pre> /mas/ic/src/VIXen/Classes/Thomas/Places/Cities.class  /mas/ic/src/VIXen/Classes/Thomas/Places/Sites.class  /mas/ic/src/VIXen/Classes/Thomas/People/Collabs.class  /mas/ic/src/VIXen/Classes/Thomas/People/Otros.class  /mas/ic/src/VIXen/Classes/Thomas/People/Researchers.class  /mas/ic/src/VIXen/Classes/Thomas/Things/Objects.class  /mas/ic/src/VIXen/Classes/Thomas/EthnoMed/Plants.class  /mas/ic/src/VIXen/Classes/Thomas/EthnoMed/Recipes.class  /mas/ic/src/VIXen/Classes/Thomas/Things/Animals.class  /mas/ic/src/VIXen/Classes/Thomas/EthnoMed/Illnesses.class  /mas/ic/src/VIXen/Classes/Thomas/Footage/Camera.class  /mas/ic/u/morgen/thesis/CLASSES/Framing.class MayaMed 334 504 3108 30 corn blowing in the wind Cities Chamula Objects corn MayaMed 505 587 3276 30 Dominga's house Cities Chamula Sites Road Objects corn MayaMed 588 700 3590 30 Laguana Pejte' Cities Chamula Objects corn MayaMed 701 1090 3590 30 dominga walks down hill Cities Chamula Collabs Dominga </pre>	 <p>HEADER</p> <p>LOCATION OF KEYWORD CLASSES</p>
--	---

The logs that were created on the Macintosh were uploaded onto the DEC station 500 workstation then translated into the Strata Data Format (SDF). The majority of the descriptions that were created in

---

<sup>1</sup>The path names for each class are delimited by “|” with only one carriage return, here they are presented one per line to facilitate legibility.

Mexico were content markers which only included a text description and an in-point. A content marker in SDF is a stratum where "In-frame," "Out-frame," and "Content- frame" are equal.

### *7.8.2 The UNIX File system*

SDF files and Keyword classes are stored as ASCII text files in directories. Editing of these files can be accomplished using conventional text editors. In addition, easy to make UNIX shell scripts can be used to parse the files of stratified descriptions. Furthermore, by having a standard data format we can load these files into the different modules.

The UNIX file system is way to structure and organize annotations and even movie sequences. Ownership can be set for access. The place where a movie is stored can provide important contextual information about the content of a sequence. This of course requires that the user is somewhat rigorous about naming and creating directories. The additional effort pays off when tracing the use of a piece of footage in the system. A consistent format for both raw footage and edited footage enables the researcher to analyze how descriptions of raw and edited footage are built up through use.

The UNIX file system provides a simple yet useful way to structure different types of knowledge about a video resource. I can allow other researchers to have access to my keyword class files by setting the permissions on the files accordingly. The last key word class in the example file shown above belongs to another user. I shared his keyword class file for framing.

### *7.8.3 Keyword Classes*

The way that keyword classes are implemented in the Stratification modules is fundamentally different then the Hypercard stack. Key words classes are organized into class hierarchies which are implemented as directory trees in UNIX. Each keyword class is stored as an ASCII text file. If desired, the user can edit the keyword class file with any UNIX text editor. Just as with free text descriptions, the choice of keywords is related to the user's intentions; they reflect the purposes and goals of a given research project.

In a multi-user environment users need to have a consistent representation of keywords for a given project. This is obvious: decisions have to be made and rules need to be established in order for a work group to code video content. These rules function as a descriptive strategy (after Wittgenstein's language games). Keyword classes help makers consistently apply keywords for a given video project. In order to find a segment of video that any particular group has coded one needs to know the rules or

strategies that were employed during the coding process. Successful perusal of the video database requires knowledge of the descriptive strategies that were used to describe the content. In the end, this consistency will help browsers. Information providers assume that the user already knows what they want to retrieve. Keyword classes provide a flexible structure that allows for consistency in naming within a particular descriptive strategy.

The games metaphor also captures another attribute of video databases. Any segment of video footage in the database points to an indeterminate number of possible interpretations - consider the changes in meaning from raw to edited/re-edited footage. On the network different descriptive strategies operate on the same set of video material (for example, one could contrast the types of descriptions employed for the purposes of discourse analysis, botanical analysis or for "Maya" home videos). In other words, a universally applicable classification system of video is an unattainable ideal. Keyword classes aid in the consistent application of descriptions *within* the scope of a project. Stratification supports different types of descriptions and allows them to coexist in the database.

The choice of keywords is related the users intentions; they reflect the purposes and goals of a given multimedia project. Keywords are only useful when used consistently. Different researchers would use different types of keyword classes that were dependent on their expertise. When conceived in this way it relieves the person who is logging the video to describe it in such a way that it can be accessible to all possible users. For example, in my log I have a keyword class called Plants.class. Figure 21 shows the Keyword classes that were employed to annotate the "Mayan Medicine" laserdisc.

Figure 21: Sample of Keyword classes for “Mayan Medicine”

```

/mas/ic/src/VIXen/Classes/Thomas/EthnoMed:
illnesses.class (symptoms, signs, duration, history)
plants.class (collecting, drying, drawing, identification, use)
recipes.class (ingredients, preparation, dose, indications)

/mas/ic/src/VIXen/Classes/Thomas/People:
collabs.class (Dominga, Xavier, Catarina, Sebastian, Alonzo, Esteban)
otros.class (children, tourists, vendors, police)
researchers.class (Brent, EAB, Luisa, John, Victor, Tere, Guadalupe, Martha, Carmelino, Domingo,
                  Feliciano, Antonio, Nicolas, Thomas)

/mas/ic/src/VIXen/Classes/Thomas/Places:
cities.class (Chamula, San Cristobol, Mexico City, Tenejapa, Cancuc, Chanaljo, Boston)
sites.class (Market, PROCOMITH, Road, Cafe, Casa, Forest, Vehicle)

/mas/ic/src/VIXen/Classes/Thomas/Things:
animals.class (chickens, pigs, sheep, birds, fish)
objects.class (shoes, hands, TV, vans, computers, video, candles, eggs, Coke, tortillas)

/mas/ic/src/VIXen/Classes/Thomas/Footage:
camera.class(panLtoR, panRtoL, zoomIN, zoomOUT, steadycam)
Transcript.class(Espanol, Tzotzil, Tzeltal, English, Japanese)

/mas/ic/u/morgen/thesis/CLASSES:
Framing.class(Extreme_Close-up, Medium_Close-up, Full_Close-up, Wide_Close-up, Close_Shot,
              Medium_Close_Shot, Medium_Shot, Medium_Full_Shot, Full_Shot)

```

The plant class contains the following list of keywords (collecting, drying, drawing, identification, use). I made this class to thematically help me keep track of plants -- I intend to edit together sequences around these keywords. Of course, a biologist could also have a class called plants but the keywords in that class could be (*Ageratina linustrina*, *Allium sativum*, *Ascaris lumbricoides* etc.). What distinguishes my keyword class “plants” from the biologist’s keyword class “plants” is location of these files in the UNIX file system. For instance, my plants class file is located in the directory /Classes/Thomas/Ethnomed. There are three files in this directory: illnesses.class, plants.class, recipes.class that reflect the different domains of ethnomedical research that I recorded. The botanist’s plant class would appear in a different directory. This directory perhaps could mimic botanical classification ontologies such as family/genus/species. With a super directory for family, an ASCII text files for each genus. These “genus” files contain entries for each species.

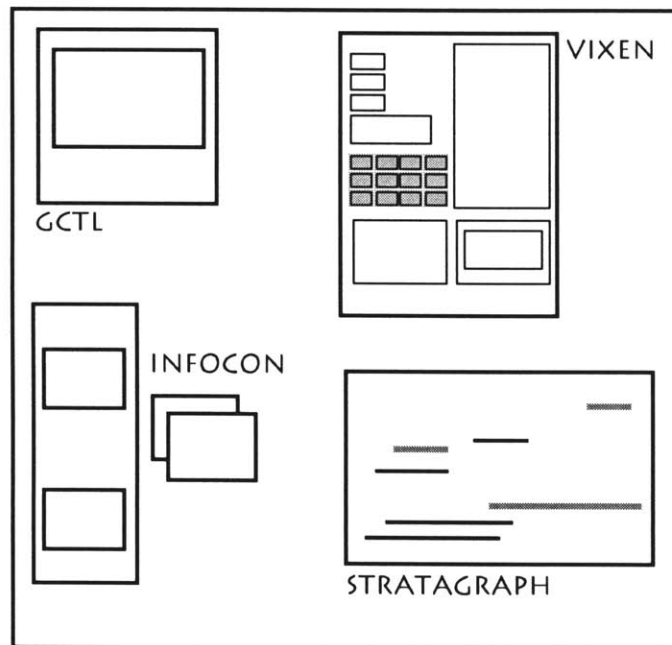


Of course, a researcher may apply more than one type of analysis to any shot of video. They might be interested in the transcript of the material in the field, or a visual domain analysis (inventories of material culture by way of scene extraction), linguistic analysis, narrative style, etc. Good source material could lend itself to be re-employed in different research environments and for different needs.

### 7.9 *The Stratification Modules*

The Stratification System consists of a set of software modules or tools sets that allow the researcher to further annotate and describe the video by using keywords (figure 22).

Figure 22: The Stratification Software Modules.



The modules include an interface to the laserdisc - "GCTL" , a video logging annotation application "VIXen<sup>1</sup>", an interactive virtual video editor "Infocon<sup>2</sup>" and a graphical display of descriptions "Stratagraph<sup>3</sup>" .

### *7.9.1 GCTL - Galatea Controller*

Logging is coordinated with browsing the video material as it is displayed in the GCTL (Galatea control) window. The GCTL window is an Xwindow that can be moved around the screen (figure 23). Using the mouse the user can fast-forward, review, search for a particular frame number and change video source volumes. GCTL also has a video slider which enables the user to quickly scroll through video in both forward and backward directions with only a click of the mouse. The slide bar allows for scaleable control of the fast forward and reverse speed of the laser disc. The farther the slider is moved from the center of the slide bar, the faster the speed of the laser disc. GCTL takes full advantage of the workstation's pressure sensitive mouse. The laserdisc is stopped within one or two frames as soon as the mouse button is released. This accuracy is essential for logging and editing.

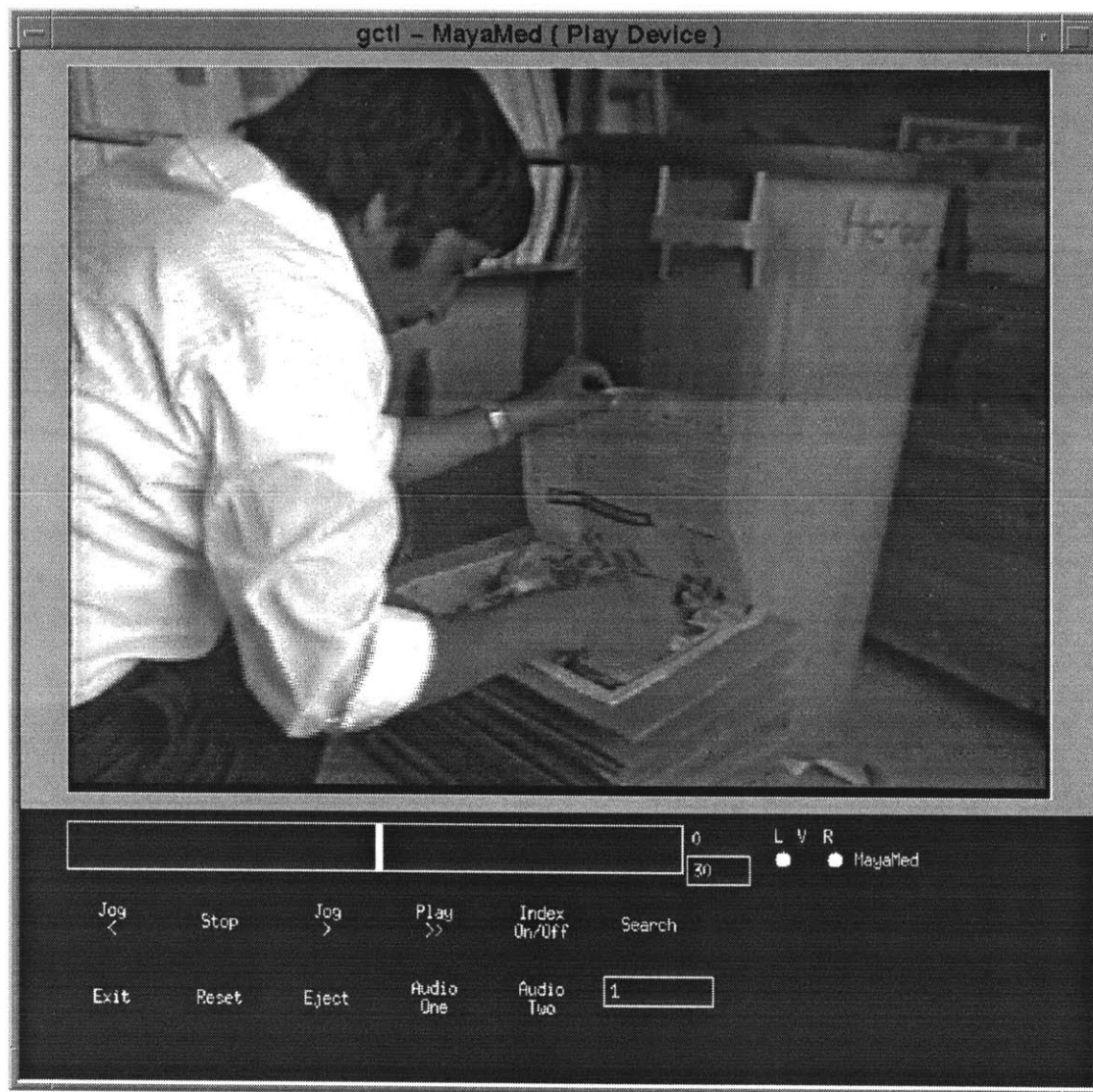
---

<sup>1</sup>VIXen was written by Joshua Holden, undergraduate research assistant in the Interactive Cinema Group.

<sup>2</sup> Infocon was written by Hiroshi Ikeda and Hiroaki Komatsu research affiliates from Asahi Broadcasting Corporation.

<sup>3</sup>Stratagraph was written by Erhhung Yuan, undergraduate research assistant in the Interactive Cinema Group.

Figure 23: The GCTL Window:



### 7.9.2 VIXen: The Logger

The VIXen is the logging and annotation module (figure 24). It is a Motif application that allows the user to enter in and out points for free text and structured keyword class descriptions. VIXen communicates with the laserdisc player via the Galatea server. Once the desired in-point / out point is found using GCTL, the user presses the in-point button and VIXen retrieves the current frame number.

Figure 24: The VIXen logging module:

**VIXen**

Volume: MayaMed  
 ID: 103  
 In: 13708  
 Out: 13782

Carmellino face

Play Open Save OK -->

Add Class Keywords Show All

Cities Sites Collabs  
 Otros Researchers Objects  
 Plants Recipes Animals  
 Illnesses Camera Framing

Medium\_Close-up

90 9535 9722 selling corn  
 91 9723 9969 selling cookies  
 92 9970 10415driving though town  
 93 10416 10525van passengers  
 94 10526 11151van stopping  
 95 11152 11571getting out shot from ir  
 96 11572 11700getting out exterior vie  
 97 11701 11888sorting the database  
 98 11869 12193walking up the stairs  
 99 12194 12323Procomith Sign  
 100 12324 12678entering and sitting dc  
 101 12679 13370BB entering room and  
 102 13371 13705orange flowers  
 104 13783 13909picking succulent  
 105 13910 14023medium shot of lupe  
 106 14024 14294placing plants in press  
 107 14295 14898chopping  
 108 14899 15102fbding straps  
 109 15103 15553saying goodbye  
 110 15554 16267placing plants in dryer

Available Keywords

Close\_Shot  
 Extreme\_Close-up  
 Full\_Close-up  
 Full\_Shot  
 Medium\_Close-up  
 Medium\_Close\_Shot  
 Medium\_Full\_Shot  
 Medium\_Shot

Use Keyword:  
 ^

<= Add Delete

VIXen annotations can be of the free text of keyword class type. The modules also supports two approaches for annotation: segmentation and stratification.

#### *7.9.2.1 Free Text Description*

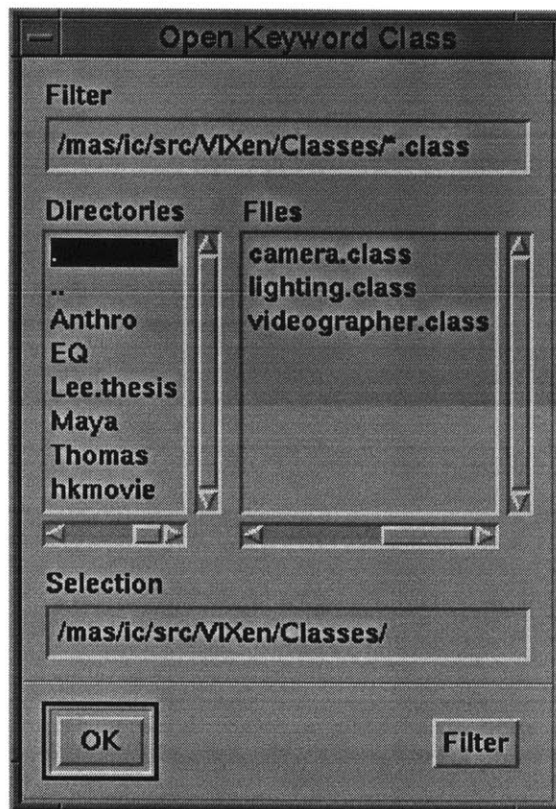
The free text fields are for free forms of description that serve as a mnemonic aid. Consider the following chunk of video as an example: Dominga whacks her grandchild as she climbs the fence that surrounds the altar. When I logged that video tape I used the word “whack” to describe the chunk. The word “whack” is stuck in my memory -- a search for this word will retrieve the shot. The strength of free text descriptions is that they reflect my personal relationship to the material and are a trace of my intentions while shooting. But there is a liability in using free text descriptions, if I forget or am not the person who made the initial descriptions then the usefulness of free text is diminished. One way around this problem is to use more generalized key word descriptions that are organized into classes.

#### *7.9.2.2 Keyword Classes*

As mentioned above, keyword classes are sets of related words which can be created, stored and be reused for different projects. For example, the “camera” keyword class ( wide shot, medium shot, pan left, pan right, zoom etc.) can be used in many different projects.

When the user opens a class file (via the “Add Class” button), a file selection dialogue box appears (figure 25). Here, the user can navigate through the UNIX directory structure to find a desired keyword class. When the desired keyword class is found and selected, the file selection window disappears and button for that class is created and displayed. Many different classes can be loaded during a logging session.

Figure 25: Dialogue for Selecting Keyword Classes



When the user wants to associate a camera keyword with a strata line, he clicks on the camera class button and all the keywords of this class are displayed in the “Available Keywords” window in the VIXen module. Double clicking on a keyword in this window associates it with that chunk of video. To disassociate a keyword, just double click on it.

### 7.9.2.3 Annotation Techniques

VIXen supports two different styles of annotation with keywords: the conventional segmentation method as well as the new Stratification method.

One can think of segmentation as a way to enter both content markers and keyword classes simultaneously. Segmentation is a brute force method. The video is annotated sequentially and a new chunk is defined for each change in descriptive state. VIXen facilitates this annotation style in a couple of ways. The out-point of the previous chunk automatically becomes the in-point of the next chunk. VIXen also keeps the list of keywords that were used in the previous segment active for the next chunk of video.

If needed keywords from the active set can be deleted or appended to reflect the changing descriptions. This feature exploits the fact that contiguously recorded footage has a set of descriptive attributes which are related to the environment where the recording took place and that these descriptive attributes remain in effect until the recording is stopped.

In figure 26, a short chunk of the laserdisc that was shot in the town of Chamula has nine of significant actions annotated. In terms of logging this footage, the user would be required to type “Chamula” for each of these chunks. VIXen facilitates entering this redundant information. The user first selects the keyword class “Places” for the first annotation. For the other annotation remains in effect. Whenever a new chunk is created the keyword “Chamula” is automatically entered. Of course new keywords could be applied or subtracted according to changes in descriptive state. VIXen allows the user to easily add redundant information which is a requirement when logging with segmentation. Of course we now have the problem of having the keyword Chamula repeated 9 times<sup>1</sup>.

Figure 26: Example of SDF Generated with the Segmentation Method.

```
/mas/ic/src/VIXen/Classes/Maya/places.class|/mas/ic/src/VIXen/Classes/Maya/people.class
MayaMed|334|504|490|30|corn blowing in the wind|places|Chamula
MayaMed|505|587|556|30|path to dominga's house|places|Chamula
MayaMed|588|700|600|30|Laguana Pejte'|places|Chamula
MayaMed|701|1090|928|30|dominga walks down hill|places|Chamula|people|Dominga
MayaMed|1091|1266|1100|30|chicken's|places|Chamula
MayaMed|1267|2041|1505|30|wacking kid|places|Chamula|people|Dominga
MayaMed|2042|2263|2130|30|pressing boys chest|places|Chamula|people|Dominga
MayaMed|2264|2483|2300|30|pressing arm|places|Chamula|people|Dominga
MayaMed|2484|2904|2779|30|throwing plants out|places|Chamula|people|Dominga
```

Segmentation can be thought of as a “brute force” method of creating a stratified description. In the diagram above we have the keyword “Chamula” in effect from frame 334 to frame 2904. The cognitive load for creating redundant descriptions can be unbearable. Furthermore, it is often difficult to keep track of everything that is going on at a particular moment. For each new observation, all the redundant descriptions from the adjacent records have to be entered.

To avoid these problems we can create stratified descriptions from the start. With VIXen we can create free text descriptions which serve as content markers (in = out = content frame) and use key-

---

<sup>1</sup>The redundancies of this method are easily handled by the Stratagraph application which is described later.

words to enter stratified descriptions. We can enter multiple strata lines simultaneously. This is best illustrated by an example. In figure 27 we have the same chunk of video that we annotated with the segmentation method. Here, each free text description is a content marker. The keywords have been entered as strata for the last three records.

Figure 27: Example of SDF Generated with the Stratification Method.

```
/mas/ic/src/VIXen/Classes/Maya/places.class|/mas/ic/src/VIXen/Classes/Maya/people.class
MayaMed|334|334|334|30|corn blowing in the wind
MayaMed|505|505|505|30|path to dominga's house
MayaMed|588|588|588|30|Laguana Pejte'
MayaMed|701|701|701|30|dominga walks down hill
MayaMed|1091|1091|1091|30|chicken's
MayaMed|1267|1267|1267|30|wacking kid
MayaMed|2042|2042|2042|30|pressing boys chest
MayaMed|2264|2264|2264|30|pressing arm
MayaMed|2484|2484|2484|30|throwing plants out
MayaMed|334|2904|1619|30| |places|Chamula
MayaMed|701|1090|895|30| |people|Dominga
MayaMed|1267|2904|2085|30| |people|Dominga
```

Each stratum was first created by associating a keyword to just an in-point. The stratum for “Chamula” was started with the in-point 334. A keyword without an end-point indicates that it is still in effect. The disc was advanced to 701 where Dominga enters the scene. Here, the keyword “Dominga” is made active. Now we have two keywords that are in effect. We advance the disc some more and find that Dominga leaves the shot. As such, we enter an end point for the “Dominga” keyword (it is selected in the list window and an end-point is entered to turn it off). The first description “Chamula” is still in effect. Dominga reenters the shot at 1267 and the camera is turned off at 2904. At frame 2904 the keyword “Chamula” is turned off. In this way, the user can easily keep track of multiple keywords which are in effect at the same time. If something needs to be added then a new strata line only needs to be entered.

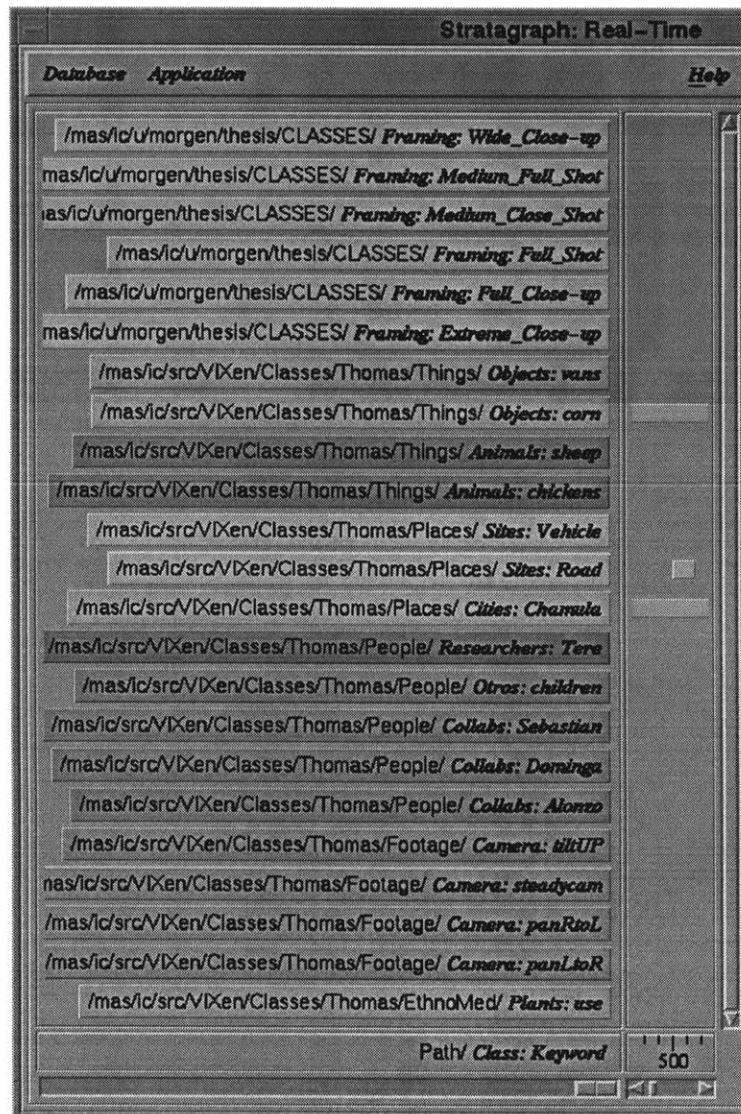
This method is a form of compression. Unlike the segmentation method, the keyword Chamula would only need to be used once - its in-point would be the first frame of the shot of contiguously recorded images and the last frame would be when the camera was turned off.

### 7.9.3 *Stratagraph*

The Stratagraph is an interactive graphical display for SDF files. It is a visual representation of the occurrence of annotations of video though time. Keyword classes are displayed as buttons along the y-axis (figure 28).



Figure 28: Key word Classes in Stratagraph



Each button shows the keywords path name in the UNIX file system. The path name indicates the context for each keyword class. On the y-axis one can inspect where a particular keyword is from and how it is related to other keywords. In this way, the UNIX file system provides a structure for organizing and representing knowledge about the descriptions of video. More sophisticated ways to represent

such knowledge such as Ken Haase's IDX system<sup>1</sup> will be implemented at a later date. In any case, there can be many types of descriptions which are generated by different ontologies. These all can be included on the y-axis. For example as shown in figure 27, the first strata displayed belong to the user "morgen", I have used his keyword class for "framing" instead making my own.

Each keyword class has its own color. The keyword classes on the vertical axis are also button widgets. When pressed, the graph scrolls to display the first instance of that keyword and the video is cued to the in-point via Galatea. To find out more about that particular instance of the keyword, the user can click on the keyword stratum and a report is generated in the "Stratum Specifications" window for that instance of the keyword. If a free text description was also associated with this stratum it is also displayed in this window.

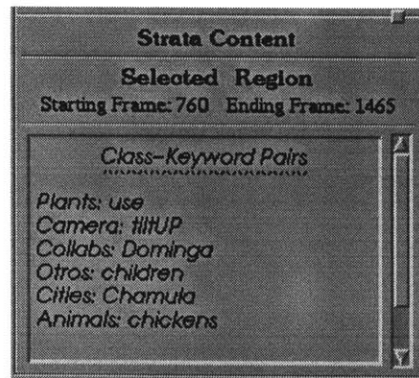
The units on the horizontal axis are time code (frame numbers for the laserdisc). Another type of interaction consists of clicking the horizontal axis with the mouse. If the user wants to know about the annotations that are associated with any particular frame. The user can click on a frame number (the horizontal axis) to create a "strata line" that intersects all the strata that are layered on top of that particular frame. The laserdisc is cued to the frame number selected and a report showing all the descriptions that are associated with these strata lines is displayed in the "Strata Content" window.

The strata line can be extended for a chunk of video by clicking the right mouse. The left click and move and right click action is called a "strata rub." This rubbing action displays all descriptions which are in effect for that chunk of video are in the "Strata Content" window while the laserdisc plays the shot (figure 29).

---

<sup>1</sup>Haase, K.(1992). IDX System Manual. (Music and Cognition Group) MIT Media Lab

Figure 29: Strata content window.



The stratagraph is an elastic representation of the content of video because any frame or chunk of video can be selected and the annotations that are associated with it can be inspected. The maker is no longer restricted into defining the units of description before hand. Furthermore, different SDF files of the same laserdisc can be simultaneously viewed in the stratagraph. The stratagraph provides an interface for analyzing how a chunk of video is described according to different lexical representation schemes or ontologies. Of course, ontologies can contest each other but in appreciating the difference and diversity of representation schemes do we attain more knowledge about the content of the video stream. We now have a indication of the most important chunks of video. The viewer, when browsing could select the chunks that have the thickest descriptions.

#### *7.9.3.1 Modifying the Display*

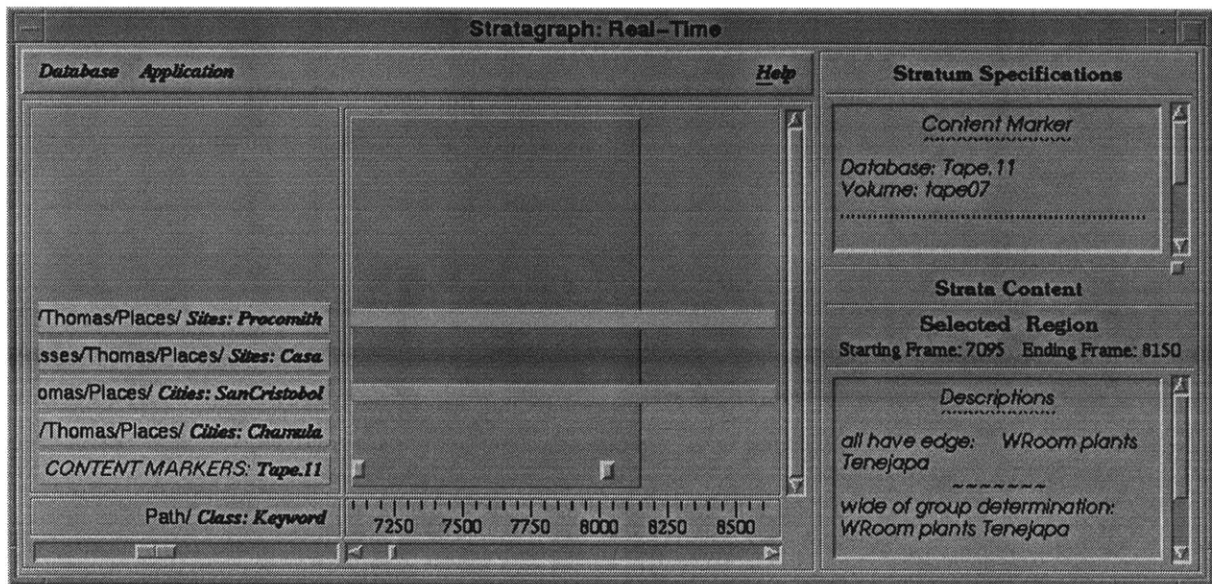
The x and y axis of the stratagraph can also be modified. Two different x-axis scales are used to display the stratagraph: the real time scale and the "scale of difference. The y-axis can "collapse" redundant keywords that have been entered using the segmentation method into strata lines.

##### *7.9.3.1.1 Real Time Display.*

For the real-time display, the scale of the x-axis is time code (frame numbers for laserdiscs). This scale is well suited for seeing the duration of a keyword -- the length of the stratum is directly related to the length of the video.

A SDF file for Tape 11 has been loaded into the stratagraph (figure 30). This file contains content markers which were created with the field notebook in addition to four keywords that have been entered as strata.

Figure 30: Stratagraph for Tape 11 showing real - time display.

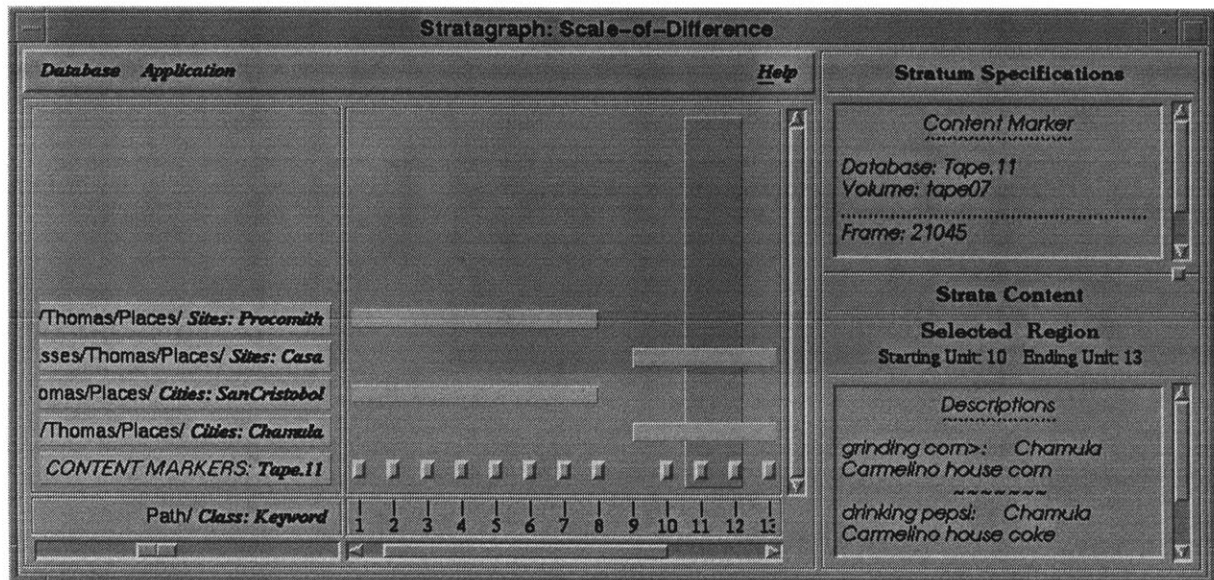


The empty space between content markers provides information about the duration of descriptions and how they are distributed throughout the tape. The empty space does not provide additional descriptive information. If I want to see the third or fourth content marker I have to scroll. If I want to jump from the first content marker to the fourth content marker (which can be done by clicking on the content marker on the graph) I will have to scroll the display. I can't get a good overview of all the descriptions for this tape. Furthermore, there is no way to easily determine when the strata lines begin or end. The "Real-Time" scale is out of phase with the scale of the descriptions.

#### 7.9.3.1.2 Scale of Difference

But sometimes it is more desirable to see where a particular description is in effect and what other descriptions are layered on top of it. The scale of difference is a compressed graphical representation of content where only changes in descriptive attributes are displayed. For the scale of difference, the coarseness of the time scale is directly related to the coarseness of the descriptions (figure 31).

Figure 31: Stratagraph for Tape 11 showing Scale of Difference Display.

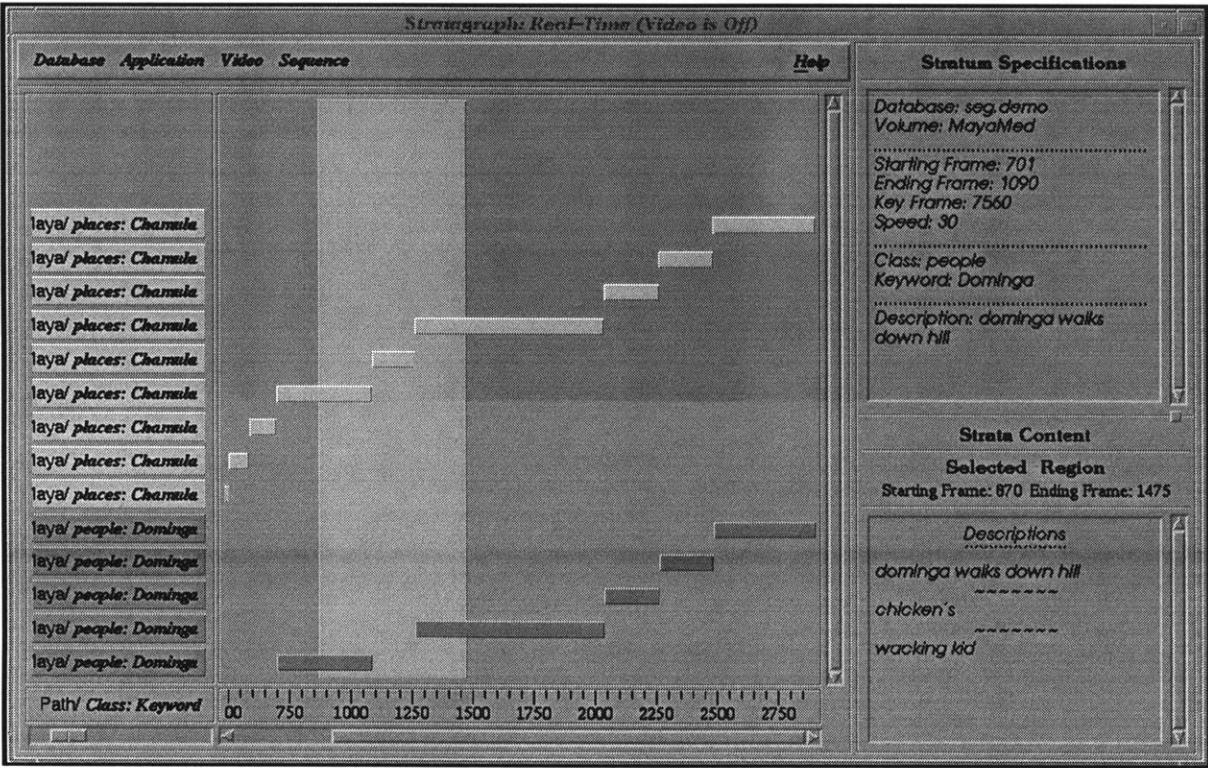


Only significant changes of descriptive state are graphed. All the in and out points are sorted and assigned a scale of difference value. For Tape 11 there are 13 unique in and out points -- the first in-point is assigned the value of 1 the second in-point gets the value 2 etc. These values are then displayed on the stratagraph. The scale of difference provides the user with a visual gestalt of all the significant actions that occur on a video tape.

#### 7.9.3.1.3 Segmentation and Collapsing Keywords

A SDF file which was generated using the segmentation method can also be displayed in the stratagraph. Each record of the file has a free text description and keywords. When plotted on the stratagraph, each record is displayed as an individual stratum (figure 32).

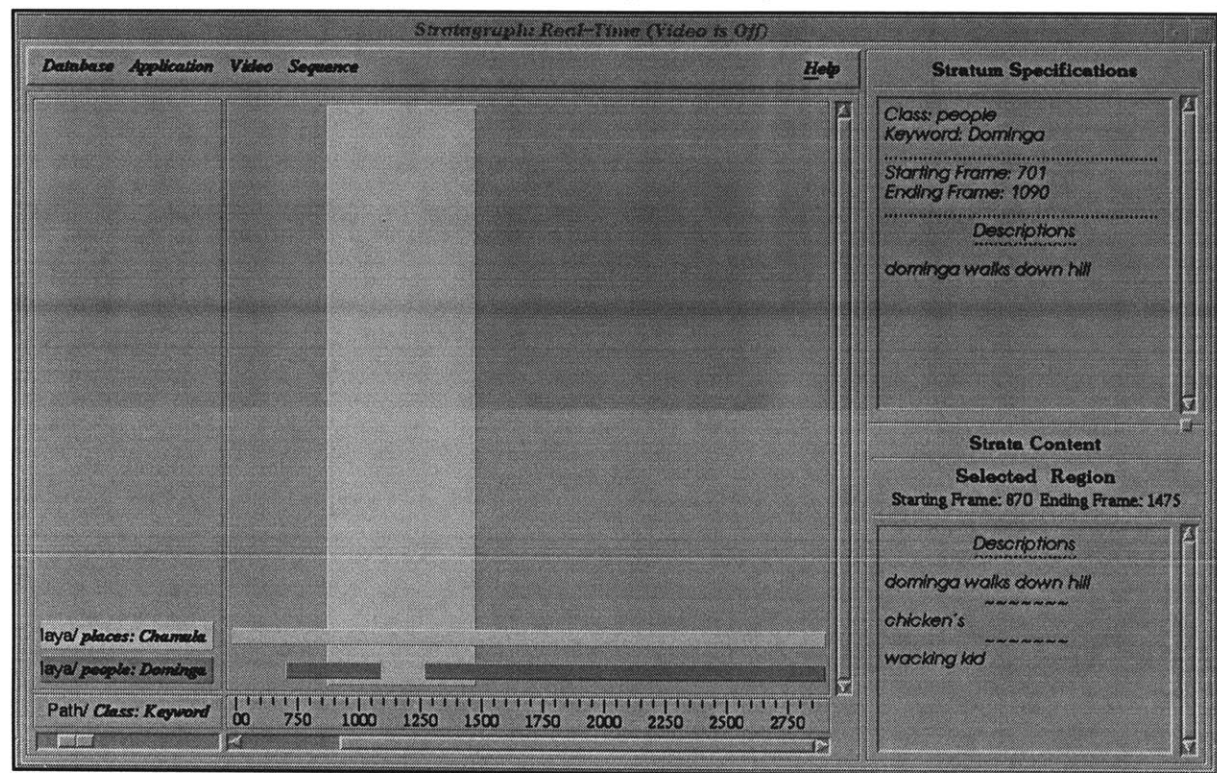
Figure 32: Stratagraph for a data file created with the Segmentation Method.



To see the free text description the user just clicks an individual strata or rubs the X axis. Unfortunately, the redundant descriptions take up too much of graph space. A segmentation SDF file that uses many different keywords would require the user to scroll up and down to see how all the different keywords are related. Another way to present this information, is to collapse the keywords into the same row (figure 33).



Figure 33: Stratagraph For the Segmentation Method--Keywords Collapsed



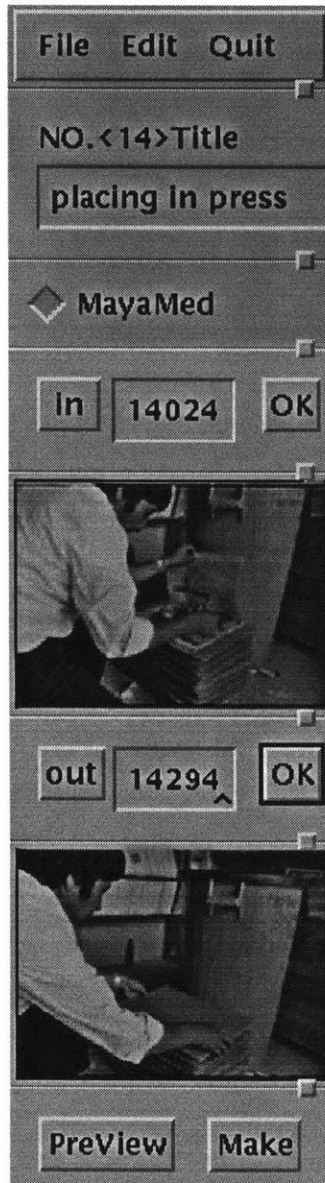
More information can be presented on the graph and the continuity of descriptive states can be readily discerned. The free text descriptions are not lost in this process. They are displayed when the user either clicks on an individual stratum or when the user rubs the graph. The user can toggle between these two modes at any moment. Strata can be generated from the keywords that have been redundantly entered with segmentation.

#### 7.9.4 Making Sequence Annotations: Infocon

To assemble a sequence the user inspects the Stratagraph for material with the desired content to select a shot. By clicking on the strata or x-axis, the video is cued in the GCTL. The shot can be further trimmed in GCTL until the exact in-point is found.

The Infocon (INfOrmation iCONinc) is the Stratification module that provides an interface for annotating shots and arranging these shots into sequences (Figure 34).

Figure 34: Infocon  
Sequence Assembly  
Module



Once an desired in-point is found, it can be quickly entered in Infocon by clicking the in button (in points can be manually entered also). The video is cued in GCTL to find an out point and it is entered. When Infocon retrieves the in- and out-frame number from Galatea, the associated frames are digitized and displayed in video windows on the Infocon module. Next the shot needs to be named and



made into an icon so that it can be arranged into sequences. When naming, the maker is creating a mnemonic that is used to reference the shot a later time. Although in and out points are important cues for visual continuity, they might not provide an adequate representation of the shot content. In addition to the in and out frame, a content frame is digitized and its associated frame number is also stored as the content marker frame for the shot.

#### 7.9.4.1 *Picture Icons - Picons*

The content frame is used to create an Picon (picture icon) which serves as a visual representation of shot content. A Picon is a free standing X window with a title bar which displays the annotation and a digitized content frame( figure 35).

Figure 35: Picons represent shots. Sets of Picons represent sequences.



Left-clicking in a Picon, plays the shot in the GCTL module. Middle clicking on a Picon displays a pop-up menu which tells which volume name, in and out points, trim button, delete, button and

a exit button to hide the pop-up. When the trim button is pressed, the PICON is loaded into the Infocon palette where the user can manually reenter the points or cue them up in GCTL.

Arrays of Picons can be displayed on the screen and arranged into sequences. When Picons are arranged in a cascade (left to right and up and down) they can be played back in the order that they appear. Ordered groups of Infocons compose a sequence. When a satisfactory sequence is arranged, annotations, volume name, in and out frame numbers are saved as a SDF file with the name of the file as the sequence name (figure 36). Sequences can be saved as play lists and can be re-loaded back into Infocon at a later time. Picons are generated on the fly each time a movie file is loaded into Infocon.

Figure 36: SDF file for a sequence "Carmelino.work.movie"

```
Filename: Carmelino.work.movie  
  
MayaMed113706113782113777130Iface  
MayaMed114024114294114026130Iarranging  
MayaMed114899115102115000130Ithe press  
MayaMed115555116260116000130Ithe dryer  
MayaMed118012118507118400130Ithe garden  
MayaMed118510119140119000130Ihouse  
MayaMed119157119353119200130Iwith healer
```

### 7.10 Naming a Sequence as Data Entry

A sequence file is a play list that shows how chunks of video are related in a new context. The sequence file provides valuable information about how chunks of video are related. The selection and ordering of chunks can imply temporal or causal relationships among chunks of video source material. The stratification system integrates this new contextual information with other annotations. Assembling sequences with Infocon is different than conventional types of editing systems because the link between source and edit is never broken. The shot that appears in an assembled sequence in Infocon is the same chunk that is logged and annotated in the VIXen module. In the stratification system there are multiple annotations or names for any chunk of source material. The first generation of annotations reflect the context of where the images were recorded, subsequent annotations reflect shifts in meaning that occur when the chunk appears in a virtual edited sequence. Logging video and assembling sequences collapse into the same process. The source material does not change per se, but the context of the material is what dynamically changes. Since cinematic context is inextricably linked to context, the significance of the source material becomes transformed and refined through use.

The path name for a SDF movie file provides contextual information for the shots that are contained in that file. The sequence file is a special case of the SDF format. When loaded into the strata graph the path name of the movie file is employed as a keyword. When displayed, the path name keyword gives the context of where the movie can be found. As with keyword classes, the UNIX directory structure can be used to semantically represent content. The way that a movie is used can be re-integrated into the Stratagraph browser. Thick descriptions arise as the chunks of video are created into shots and used in sequences. The stratagraph is used to graphically represent both logging information and annotated virtual sequences. The user can interact with these sequence annotations on the stratagraph. In this way, the virtual content of a sequence can serve as a gateway back into the source material. As the result of these explorations other sequences can be created and integrated into the stratagraph. The stratification system captures how the significance of an image dynamically changes through use and allows users to use this information to construct new meanings.

## 8. CONCLUSION

The research presented in this thesis illustrates how the integration of text and video on digital computers can transform ethnographic research. The stratification systems provides a robust representation of audio visual descriptions as they dynamically change within time and as they are dynamically transformed over time. The stratification system provides computational support for the practice of interpretive anthropology. The Anthropologist's Video Notebook is the new medium for video ethnography where the significance of the content of the video stream emerges over time as anthropologists develop their memories of observed events and utterances. The goal is to *communicate* the significance of what occurred to others.

Audio visual descriptions become inscribed at the moment they are recorded on the video tape. Although the video stream is an artifact of the contextual factors that were in effect when it was recorded, in no way is it something dead. The process of inscription is the birth of a percept and as such, it is alive for a host of possible interpretations and may appear in an indeterminate number of contexts. The value of an image on the stratification system can be measured and traced by its appearance in a variety of contexts. As the video stream percolates in the stratification system, value is added as different researchers use it to articulate their interpretations of what happened. The power of the environment lies in the relative strength of the different descriptive strategies that have been applied to any given chunk of footage.

I predict that ethnographers will not produce movies per se but will rely on the publication of video databases where lexical descriptions as well as the ordering of sequences produce thick descriptions. Descriptions in a video database are thick in the domain of their design - how they have been assembled or edited thickly (the virtual context where they are used) together with the more textual practice of lexical description which allow for the integration of new insights.

Movie production involves the creation of audio visual artifacts; video ethnography is the practice of studying what meanings people assign to these audio visual artifacts. When the video is placed in the design environment of stratification it becomes a node for diverse types of understanding and contested meanings. It becomes enmeshed in a social fabric of interpretation. The meanings associated with the video stream form a culture. Spradley (1980) defines culture as:

The acquired knowledge people use to interpret experience and generate behavior. By identifying cultural knowledge as fundamental, we have merely shifted the emphasis from behavior and artifacts to their *meaning*. The ethnographer observes behavior but goes beyond it to inquire about the meaning of that behavior. The ethnographer sees artifacts and natural objects but

goes beyond them to discover what meanings people assign to these objects (Spradley, 1980 : 7 -8).

The stratification system is a cultural space where we can explore the meanings that people assign to a video stream. With the stratification system we can begin to look at how a culture of the moving image is created and investigate the relationships between the institutions and actors in the continual re-creation of the content of the video stream and the symbols and meanings which are attributed to the video stream over time.

Stratification is an integrated representation for the content of a video stream. It maintains the descriptive integrity of the source material while supporting the re-juxtaposition of the “raw video” in to new edited contexts. There is an interplay between the rigorous analysis of video data and the maintenance of the original descriptions. The former is a prerequisite to the generation of new theories and a deeper understanding about a culture while the latter allows for descriptive coherency through out the production process - from shooting to sequence assembly.

Knowledge about the video stream has a scope that is related to the needs and intentions of the movie maker. Lexical descriptions of content are useful when considered in the context of a particular production. And as Wittgenstein has pointed out, the meanings associated with these lexical descriptions can be understood from their use along with an explanation of how they are used within the scope of a particular project. It is clear that movie makers do not want to be hindered by the task of describing the content of the video stream for every possible use. The methodology presented in this thesis allows the makers to describe video within the scope of their individualized production needs. When these esoteric descriptions are combined with other descriptive strategies for the same video resource knowledge about the content of the video stream is built. As more and more people annotate footage a memory in the present is created for the content of the video stream.

Stratification provides a way to describe every single frame and every single pixel that composes a frame. The content is discerned through the examination of contextual information as it is represented in the strata lines. It is an elastic representation of content because the content of any set of frames can be derived by examining the strata that the set of frames is embedded in. The overlapping strata lines can contest one another. At first glance there is no way to decree the definitive truth about chunk of video. But on second glance we see that stratification is a computational representation for the ambiguity of a moving image. The ambiguity of an image, especially when perceived in an edited context, is a hallmark of the expressiveness of the medium of images. It is a reflection of the subtlety and elegance of video as a form of communication. This expressiveness and elegance can now be supported on a computer system and become an integral part of the interpretive practice of video ethnography.

## REFERENCES

- Abbate, M. and P. Palombo (1991 - 92). Video Tool kit. Norfolk, MA: Abbate Video Consultants.
- Applebaum, D. (1990). The Galatea Network Video Device Control System. (Interactive Cinema Group, Working Paper), MIT Media Lab.
- Aguierre Smith, T. (1990). Negotiated Crossings: An Ethnographic Video and Research Report. (Research Theory and Methods in Anthropology) University of California, Berkeley.
- Aguierre Smith, T. (1991). Stratification: Toward A Computer Representation of the Moving Image. (Interactive Cinema Group, Working Paper) MIT Media Lab.
- Bamberger, J., & Schön, D. (1982). Notes From Work in Progress. (Department of Architecture) MIT.
- Bellman, B., & Jules-Rosette, B. (1971). A Paradigm for Looking: Cross- Cultural Research with Visual Media. Norwood, New Jersey: Ablex Publishing Corporation.
- Berlin, Brent, Elois Ann Berlin, Dennis E. Breedlove, Thomas O. Duncan, Victor Jara Astorga, and Robert M. Laughlin (1988). Medical Ethnobotany in the Highlands of Chiapas: A Progress Report on the First Fifteen Months of Fieldwork. Invited Paper at the symposium, "Plants and People: Current Investigations in Ethnobotany." Annual meeting of the American Anthropological Association, Phoenix, Arizona, 16 November.
- Clifford, J. (1990). Notes on (Field)notes. In R. Sanjek (Eds.), Fieldnotes: The Makings of Anthropology (pp. 47). Ithaca: Cornell University Press.
- Clifford, J., & Marcus, G. E. (Ed.). (1986). Writing Culture. Berkeley: University of California Press.
- Collier, J. (1986). Visual Anthropology. Albuquerque: University of New Mexico Press.
- Davenport, G. (August 19, 1987). New Orleans in Transition, 1983 -1986: The Interactive Delivery of a Cinematic Case Study. The International Congress for Design Planning and Theory. Boston : Park Plaza Hotel
- Davenport, G.,Aguierre Smith, T., & Pincever, N. (1991). Cinematic Primitives for Multimedia: Toward a more profound intersection of cinematic knowledge and computer science representation. IEEE Computer Graphics and Applications(July).

- Elliot, E. (1992). Multiple Views of Digital Video. (Interactive Cinema Group, Working Paper) MIT Media Lab.
- Geertz, C. (1973). The Interpretation of Cultures. New York: Basic Books.
- Goldman Segall, R. (1990). Learning Constellations: A Multimedia Ethnographic Research Environment Using Video Technology for Exploring Children's Thinking. Ph.D., Media Arts and Sciences, MIT, August 1990.
- Krebs, S. (1975). The Film Elicitation Technique. In P. Hockings (Eds.), Principles of Visual Anthropology Paris: Mouton Publishers.
- Leacock, R. (1986). Personal Thoughts and Prejudices about the Documentary. Presented at the 23rd CILECT Congress (November). Paris : CILECT.
- Leacock, R. (1990). On Working with Robert and Frances Flaherty. Presented on April 26. Cannes : Cannes Film Festival.
- Leacock, R. (1991). Life on the Other side of the Moon. (Draft) Paris.
- Levaco, R. (Ed.). (1974). Kuleshov on Film. Writings by Lev Kuleshov. Berkeley: University of California Press.
- Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: W. H. Freeman.
- McLaughlin, J. (1990). "Cape Cod Dig Full of Promise: Site Ancient, Time Short" Boston Globe: December 8 , front page.
- Ottenberg, S. (1990). Thirty Years of Fieldnotes: Changing Relationships to the Text. In R. Sanjek (Eds.), Fieldnotes: The Makings of Anthropology (pp. 139). Ithaca: Cornell University Press.
- Pincever, N. (1991). If you could see what I hear: Editing assistance through cinematic parsing. MS Thesis, Media Arts and Sciences, MIT, June 1991
- Rubin, B. (1989). Technical Considerations for Virtual Video Editing. (Film Video, Working Paper) MIT Media Lab.
- Sasnett, R. (1986). Reconfigurable Video. MSVS Thesis in Architecture, MIT, January 1986.
- Sanjek, R. (Ed.). (1990). Fieldnotes: The Makings of Anthropology. Ithaca: Cornell University Press.
- Schön, D. (1987). Educating the reflective practitioner. San Francisco: Jossey-Bass.
- Schön, D. (1988). Designing: Rules, types and worlds. Design Studies, 9(3), 181-190.
- Schön, D., & Wiggins, G. (1988). Kinds of Seeing and Their Functions in Designing. MIT.

- Spradley, J. (1980). Participant Observation. New York, Holt, Rinehart and Winston.
- Teodosio, L. (1992) Salient Stills. MS Thesis, Media Arts and Sciences, MIT, January 1992.
- Turk, M (1991) Interactive-Time Vision: Face Recognition as a Visual Behavior. Ph.D. Thesis, Media Arts and Sciences, MIT, September 1991.
- Wittgenstein, L. (1958). The Philosophical Investigations. New York: Macmillian Publishing Co.
- Worth, S., & Adair, J. (1972). Through Navajo Eyes: An exploration in film communication and anthropology. Bloomington: Indiana University Press.