The Use of Multimedia Technology in Distance Learning[1]

T.D.C. Little and D. Venkatesh

Multimedia Communications Laboratory
Department of Electrical, Computer and Systems Engineering
Boston University, Boston, Massachusetts 02215, USA
(617) 353-9877, (617) 353-6440 fax
tdcl@bu.edu

MCL Technical Report 08-10-1995

Abstract–Distance learning describes the delivery of instruction to students beyond the geographical constraints of a single classroom. It is commonly deployed using conventional analog video broadcast or closed-circuit technology. Advances in personal workstations and low-cost desktop videoconferencing are making all-digital distance learning a viable alternative.

In this paper we describe recent developments in multimedia computing and communications technology that can be applied to distance learning. In particular, we consider the use of both asynchronous communications (e.g., the World Wide Web and electronic mail) and synchronous communications (e.g., the multicast backbone – MBone, videoconferencing, and the use of shared whiteboards). We also discuss the design of our distance learning testbed that supports access to recorded instructional video.

Keywords: distance learning, multimedia, course-on-demand, video-based-instruction.

[1]In *Proc. IEEE Intl. Conf. on Multimedia and Networking*, Aizu-Wakamatsu, Japan, September 1995, pp. 3-17. This work is supported in part by EMC Corporation and the National Science Foundation under Grant Nos. IRI-9211165 and IRI-9502702.

1 Introduction

There will be fundamental changes in the techniques employed to educate students. These changes will occur as teachers learn how computer-based tools can satisfy existing pedagogical theory and the evolving needs of our increasingly technological society. For example, the use of video-conferencing in distance learning has enabled a new class of students to benefit from the educational infrastructure at Boston University without a significant disruption in their job activities. More profound changes are anticipated as educators recognize the benefits and limitations of the technology. Developments in communications and computing, and the affordability of information delivery services will further increase this impact. By 2010, education will be very different from today. Content will still be paramount, but the tools and techniques for instruction will change. In this paper we focus primarily on distance learning, but we believe these techniques will become integral with basic educational technology.

Historically, distance learning describes a form of education in which students interact with an instructor who is located beyond the confines of a conventional classroom. A correspondence course via post is an example of distance learning. Students are mailed course material on a periodic basis. The exercises must be mailed back to the instructor who in turn grades and returns the material. Any interaction between the student and the instructor is limited to that conveyed by written correspondence or, occasionly, by telephone. Often it is difficult for the student to gain access to additional supporting course materials. In summary, the main drawbacks of such a scheme are the latencies of communication, limited interaction, and limited access to supporting materials.

Television and radio broadcast technologies have vastly improved the correspondence course model. With these technologies, a large number of students can receive instruction in real-time and with the added impact of aural and visual media. However, interaction with the instructor is still limited by the communications “backchannel,” i.e., the postal service or telephone. The use of facsimile and electronic mail introduce a form of asynchronous communication with the instructor that reduces latency considerably. All of these techniques have been used extensively in higher education.

Course materials in conventional distance learning are usually more thoroughly prepared rather than delivered on-the-fly notes. This is due to the need for conveying the same information accurately to many students who may not be able to react quickly to any problems discovered with the materials. The limitations of the technology also dictate the preparedness of the instructor. For example, visuals are more clearly viewed by television when they have been produced as type instead of handwriting. Unfortunately, this rigidity can compromise the spontaneity of instruction.

Evolving multimedia computer and communications technology are poised to lift these limitations. Course content can be created and posted for distribution via the World Wide Web (WWW); audio and video can be multicast to students throughout the Internet; and students can communicate with instructors though low-cost videoconferencing hardware. Fig. 1 illustrates this scenario.

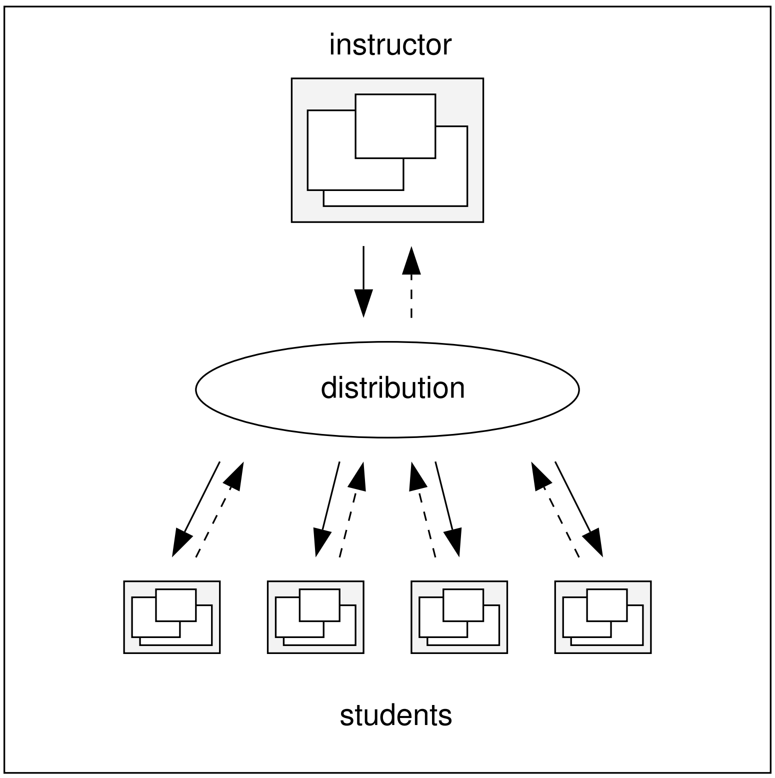

Figure 1: Electronic Distance Learning

As we move to this all-digital universe of information there are some fundamental unanswered technical questions to be resolved before ubiquitous digital distance learning is viable. For example: How can existing videoconferencing services be scaled to support hundreds of of students with low-cost and reliable reception? How can information creation, collection, and indexing be simplified to support straightforward student access? How can a very large information space be searched and filtered based on individual student needs?

Our primary focus in this paper is the design of a distance learning prototype that has the ability to quickly access information encapsulated in large video data sets. This need is illustrated by the following scenario.

A Distance Learning Scenario: Consider a real application of distance learning at Boston University. Engineering classes are held in one of two broadcast studios which transit live video to approximately 40 sites off campus. Each course typically produces 3-4 hours of video per week for 13 weeks. The video is recorded on video tape as it is transmitted. Possible uses for this recorded material include the following: a student misses a class and needs to replay the class at a convenient time, wishes to review a particular topic in preparation for an exam, seeks help on a subject overlapping a previous course, or seeks all solutions to problems sets described in previous years.

For a number of reasons, the video tapes do not satisfy these needs and are of little use after they are produced. Here is why:

1. The information-to-data ratio is very low. Often, an instructor will show typed "transparencies" that are transformed from simple text to lengthy video sequences with brief pointer movement (Fig. 2).
2. Lengthy video tapes of lectures are not "riveting" and are best edited.
3. It is difficult and time consuming to locate and use the information encapsulated in the medium.

The last property is due to the sequential nature of the tapes. To locate a particular topic within the course, one must know which tape (out of 26) to select, and at what point on the tape the topic is covered. By simply digitizing the video tapes, one can achieve random access to the material but without the necessary indices to locate sought topics.

For this application domain there are potential solutions which we intend to investigate that ultimately lead to "fast access" of the content. This will require transformations on the digitized video stream. For example, to eliminate redundant data, only unusual events need to be captured. These events include gestures, expressions, and other special events by the class or instructor. These "events" can also represent information content changes such as the introduction of new transparencies. Clearly there is a need for domain-specific information models. These models should be designed to permit extensive cross-referencing and indexing of the variously collected data: audio, captured transparency stills, gestures

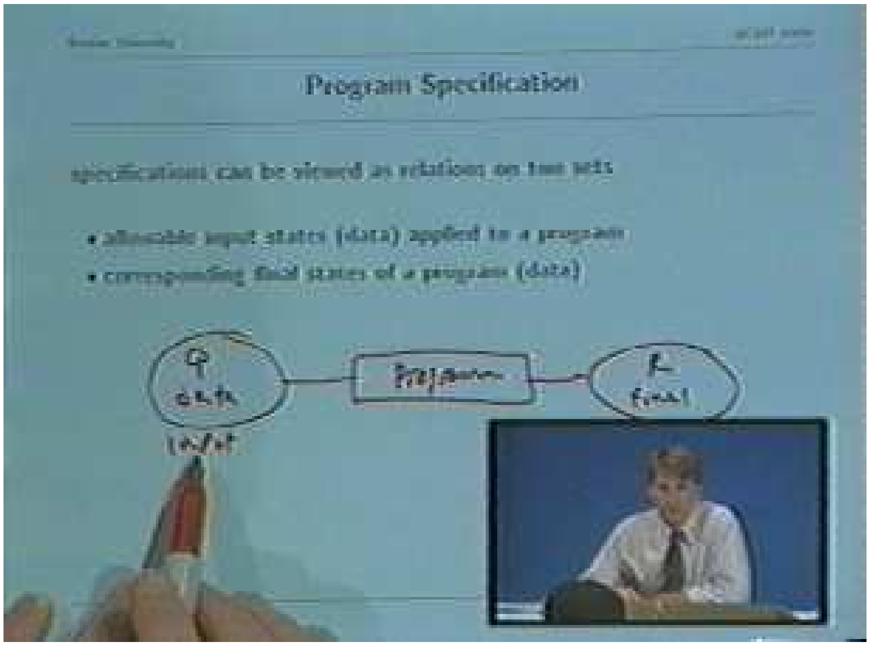

Figure 2: Video-Based Instructional Content

(including pointer movement), and text-based materials. Finally, these need to be organized in a way that economizes both the access time for the student and the transmission/network bandwidth required by the information system. Similar scenarios exist for entertainment and information (news) video delivery, but with different content models [14].

We envision a system that will facilitate electronically mediated classroom instruction in this context as well as *personal* course delivery to the desks of students without the need for centralized receiving stations. A student would be able to participate in class from the office or home, or from any city, through network interconnectivity. The proposed system would allow wide-area accessibility to course instruction and would facilitate deferred, self-paced instruction when coupled with a data/information storage facility.

By using a computer-based solution, new and more efficient means of communicating among students and the instructor are possible. Homework problems could be assigned and submitted electronically though the same delivery mechanism, as might examinations (e.g., via electronic mail or other protocols). Furthermore, computer group-mediation tools could be used to manage discussion, and to prevent a "party line" atmosphere. Ultimately, such a system will support on-demand replay of any lesson at any time, thereby allowing students to go back to review concepts at their personal convenience.

The remainder of this paper describes the evolving technologies necessary to support distance learning. In Section 2 we identify the technological issues that must be resolved in building a distance learning system. In Section 3 we discuss some existing prototypes. In Section 4 we discuss the architecture of the proposed Saddle testbed. Section 5 concludes the paper.

2 Requirements for Distance Learning

In this section we identify the various technological issues that must be addressed in the construction of an all-digital distance learning environment. This includes the satisfaction of basic functional requirements and the needs dictated by communication, data collection, and storage.

2.1 Functional Requirements

Of primary importance in computer-based education is the view that the computer is a tool to help achieve an objective. In this context it is a tool to enhance learning under geographical and time scheduling constraints. Specifically, it can enable information distribution, reuse, and organization. It also enables communication and collaboration between the instructor and student, and among students (Fig. 3).

Three modes of instruction can be defined: *live* or *synchronous* instruction, in which an instructor lectures or interacts with a set of students in real-time; *deferred*, *asynchronous*, or *self-paced* instruction, in which students interact with course materials that are usually prefabricated; and *hybrid* techniques that combine both of the former. Live instruction originates from a central facility (e.g., campus) which can use more extensive production facilities including multiple cameras and staff. Prefabricated instruction can be prepared as would any instructional material, and be stored in digital format on a storage server (e.g., the WWW).

Basic functional requirements for distance learning are the ability to distribute live instruction, create prefabricated, *static* instructional materials, and to rapidly convert live materials to static ones. Additional requirements include capabilities to moderate interaction in the virtual classroom (i.e., floor control), submit assignments to the instructor, and for the instructor to respond to students both synchronously and asynchronously. The sys-

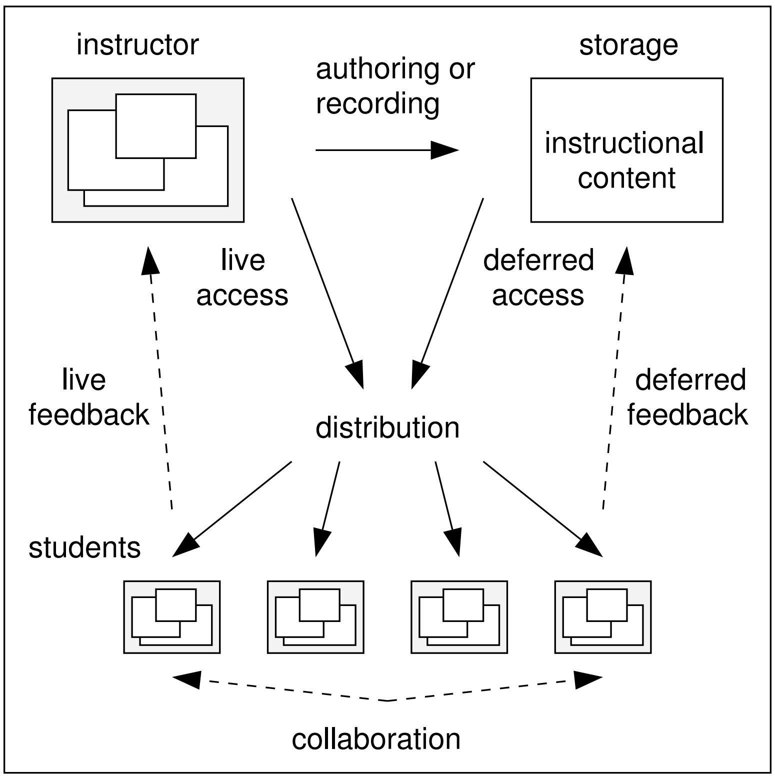

Figure 3: Communication Paths in Distance Learning

tem must also provide a high degree of communication among participants (students and instructors), and the functions to retrieve content rapidly.

Important application software issues include the development of appropriate mechanisms for floor control during lectures, multiplexing continuous audio with video stills (or faxes), video indexing, information location and retrieval, broadcast of video within the network (protocols), and connection failure recovery. Additional interfaces for authoring, inputing queries, graphical visualization of content, and mechanisms for browsing, storing, are also necessary.

With these functional requirements in mind, we consider more detailed requirements that must be addressed by the system.

2.2 Organization of Instructional Content

A very important aspect of any envisioned distance learning system is the ability to rapidly create, organize, store, and retrieve a great deal of multimedia content including audio and video. These tasks require adequate editing tools for content synthesis (e.g., word processors, drawing tools), content capture (e.g., audio recorders/editors, video recorders/editors), content organization (e.g., database management systems, hypertext editors), and content viewing (e.g., WWW servers and viewers). Of particular interest in the distance learning environment is the need to find subject matter in vast amounts of video content.

As described earlier, access to topics found on video tapes is is grossly deficient due to the sequential storage medium. By storing video data on a random access medium such as magnetic disk, we alleviate this access problem provided that a suitable set of indices are created for a given instructional segment. Given the availability of these indices, it is also possible to tailor instruction to a student's level of comprehension. For example, an instructor will often describe supporting examples to illustrate a concept presented as theory during a lecture. This content can be indexed, segmented, and ultimately retrieved in a recorded sequence based on a students needs. Fig. 4 illustrates such a scenario in which the individual topics are represented as instructional units containing video.

Important requirements to support this vision are the abilities to create the indices and locate video objects.

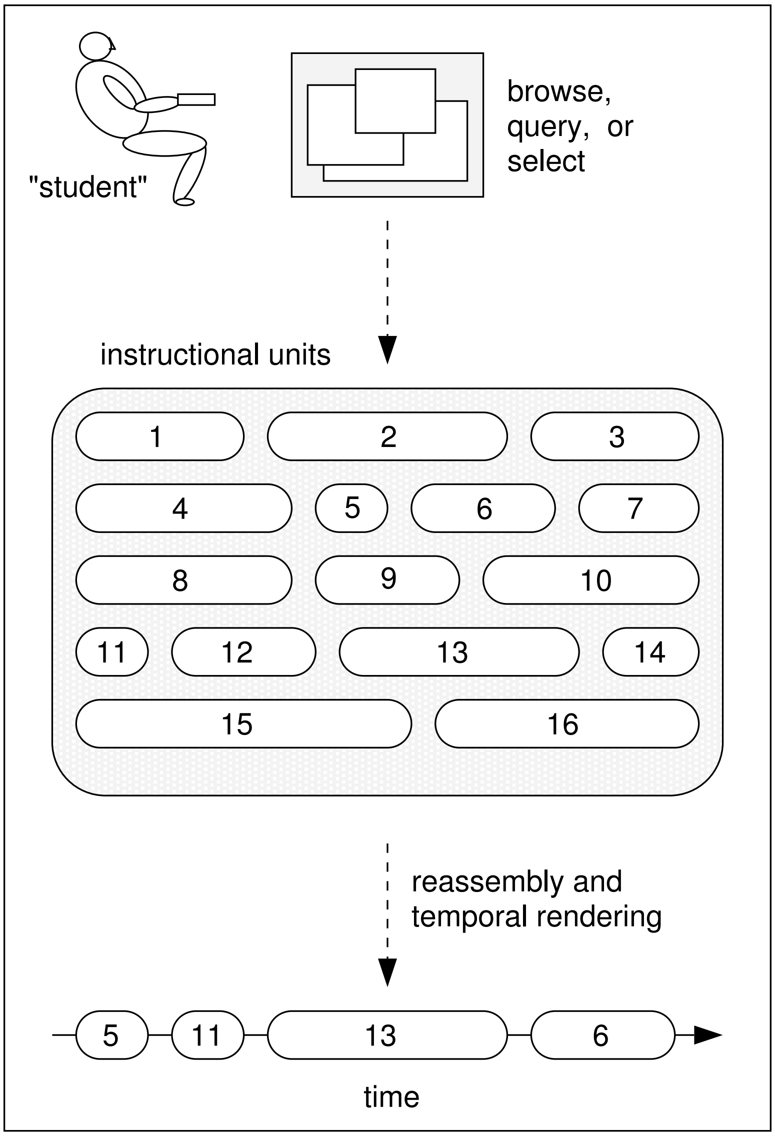

Figure 4: Selection Process for Instructional Units

Indexing of Video Content: Creation of the video objects can be achieved by extracting video segments from instruction originating from a recorded lecture, or can be created and edited independently from the live course. In the former, tools are necessary to facilitate rapid conversion from live instruction to recorded and indexed topics, and the linking of related static course materials (e.g., references to sources and homework sets). These same tools should also allow the elimination of out-takes, or other errors that the instructor/editor deems unacceptable. Significant data compression is possible in this step. Materials created in electronic form, but displayed and recorded as video in a lecture can be associated with the audio soundtrack of the selection. For example, a still graphic produced as digital line art and text, but recorded on video during presentation can be substituted in place of video during the editing process. In this sense the original audio/video recording of a lecture represents a serialization of the original content, as necessitated by the transmission medium. By returning to the original source materials in their native digital formats, one can achieve significant data compression over the recorded video stream, and can readily achieve topic decomposition for indexing the audio/video segments.

Tools to facilitate the decomposition and indexing of the stream should be designed to simplify postprocessing of the data stream for rapid conversion to static course materials and, if at all possible, eliminate the need for two passes over the data. For example, because the capture of live streams is sequential (i.e., it takes one hour to capture one hour of a live event) there is an opportunity to take advantage of this time to introduce indexing information. Moreover, we argue that the indexing process for the recorded video is best achieved by the the domain expert: the instructor who created the content. It is easy to imagine a scenario where the recorder has information or knowledge which would allow the recognition of an important event during the capture of a stream; an editor lacking the same information may not understand the significance of the event. Using an example of a baseball game, an event might be "player steals a base," setting the all-time record for bases stolen in a single season. To the announcer and/or recorder of the game, as well as to future viewers of the video stream, this represents a major event. To an editor preparing this stream for viewing who happens to not follow baseball, the event may just be another stolen base and the event may not be indexed. This is an example of information loss due to an editor unfamiliar with the domain of the stream which is being indexed.

Metadata Management: Once the video data are indexed, there is a data management and access problem. For example, if the "instructional units" of Fig. 4 consist of complex multimedia objects, there is a need to locate their components, assemble them, and deliver them in a timely manner to the student. Moreover, because video data are best served by a storage system supporting continuous media, it is desirable to separate the data searching and location functions from the raw storage functions. For these reasons a distinct meta-data management scheme is appropriate for multimedia instructional content. Appropriate search tools and engines are also needed. These tools can be conventional text-based engines assuming the use of metadata. More sophisticated image-based or motion-based tools for content searching can be used, but their utility in this application domain is not clear at this time.

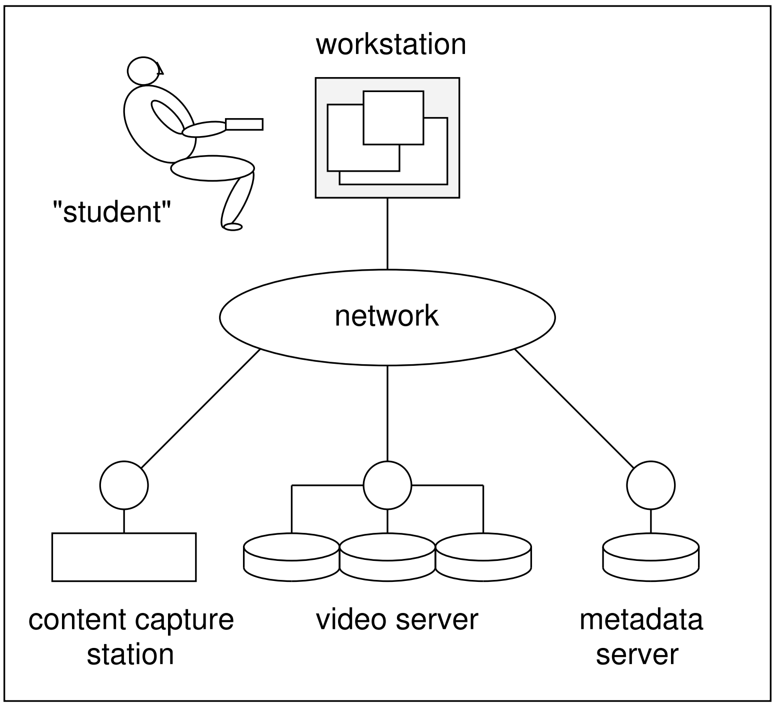

Figure 5: System Model for Distribution of Instruction

2.3 Hardware Components

The functional needs dictate the requirements for computer and communications infrastructure. The principal ones are the support of aural and visual instruction via live video, and the asynchronous delivery of prefabricated course materials. Video instruction can successfully be achieved without a live “backchannel;” however, we describe a system that can support this bidirectional communication. The components of such a system include a video capture station in the classroom, student workstations, a network switching function, and a storage server (Fig. 5).

Capture and Student Stations: The instructor's video capture station can be of high-quality because it is a single component that does not need to be replicated. In analog video distance learning deployments, this function is satisfied by multiple high-quality video cameras, microphones, video production facilities, and staff. In contrast, the student work-stations must be inexpensive and the overall system must support a variety of terminals and terminal characteristics.

Such a system will need low-cost terminal equipment such as PCs that can support digital video capture compression/decompression and display. Current prices for such a package are in the $4–5K range; however, interoperability among platforms is not widespread. Note that many of these systems are not optimized for the transmission of video images of handwriting or printed text.

Multicasting: The system must also support a multicasting or "bridging" function that allows multiple participants to view the same data stream. In conventional video broadcast systems, this function is satisfied by virtue of the broadcast medium. In switched digital communications, this function must be provided elsewhere. It can be provided by a local hardware device called a multipoint control unit (MCU), by a network service provider's MCU, or by virtue of the network's routing protocols (e.g., via `tunneling` in the MBone [5]). The provision of the MCU function is probably the biggest inhibitor to distance learning through interactive video: the support of tens or hundreds of simultaneously connected students requires cascading multiple MCUs at significant cost.[2] The MBone (multicast backbone), a virtual subset of the Internet, can support the multipoint function for such a large number of participants, but cannot guarantee available network bandwidth necessary as a foundation for an educational program.

Communications and Networking: Computer networking and communications are fundamental to distance learning. Existing telecommunications technology is capable of satisfying all communications requirements outlined in this paper, but often at significant cost. However, the rapid changes in technology deployment are likely to bring the cost of some of the technologies into the realm of distance learning.

[2]Point-to-point videoconferencing does not require the use of a MCU.

Evolving information delivery applications have already become the dominant bandwidth consumers on the Internet and are expected to be so on future CATV (common antenna television) and PSTN (public switched telephone networks) along with video-on-demand (VOD) services. This new kind of data traffic is a departure from that of electronic mail and bulk file transfer typically supported by "data" networks; that of broadcast audio/video via CATV; and of low-bandwidth audio supported by PSTN networks.

From a technical perspective, these changes are being accommodated by addressing the shortcomings of each communications infrastructure. On the packet-switched Internet, mechanisms to support real-time, continuous media (CM) traffic are being developed. For CATV, the need for bidirectional signaling is being addressed to support interactive programming and information retrieval. For the PSTN, the need for higher bandwidth and the ability to support multipoint applications is considered. In all cases there is a trend towards supporting mixed bidirectional services. However, there is a fundamental tradeoff between individualized service and a large receiver population. Broadcast technologies (e.g., CATV) must be modified to support individual interaction by data recipients. Point-to-point technologies (e.g., Internet, PSTN) must be adapted to support multipoint data distribution. Fig. 6 illustrates such a spectrum of possible services for interactive video delivery. Case (a) requires a high-bandwidth outbound video stream from a data server and a low bandwidth interactive return signal. Case (b) is the multipoint scenario without interaction. Case (c) is fully interactive wherein each client requires a point-to-point session. This final case consumes the most system resources, but is typical of what we need for interactive distance learning.

For asynchronous distance learning, connections tend to be point-to-point between the student and the content server. In this case the multipoint function is irrelevant, and the system is required to deal with the aggregate access bandwidth demands of the asynchronous student population. For conventional instructional content delivered over the Internet (devoid of video), data transfer is due to small, well-defined units such as text documents or still images. In this mode, very large numbers of users can be supported by a single server because the transmission load is intrinsicly multiplexed on these units. For audio and video data, the same servers must dedicate significant resources (access bandwidth) to satisfy a single student. Moreover, often the transferred video clip is found to be the wrong one sought due to inadequate description of its content. A solution proposed is to appropriately fragment the video material for economical exchange of information.

This leads to two complementary networking requirements: (1) the ability to simultaneously service a large number of students/connections via modest bandwidth per connection with the same data stream, and (2) the ability to service many video streams from a server. The implications of (1) are the need for many ports for connecting students to the system and the need for multicasting. The implication of (2) is the need for a large storage-to-port bandwidth based on the number of asynchronous students supported at any one time.

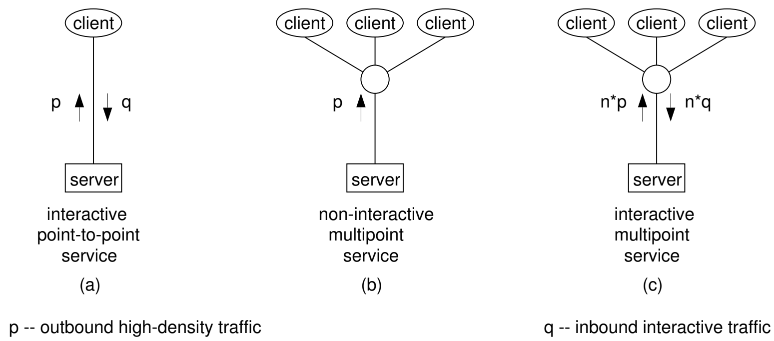

Figure 6: A Range of Interactivity for Distance Learning

Raw Data Storage: An important aspect of education is the accessibility of instructional materials. This refers to books, syllabi, homework assignments, etc. In distance learning, this also includes audio and video recordings of class instruction. Current strategies for placing video taped instruction "on reserve" in a library for student access are functional but provide poor access to specific content. With the exception of copyrighted materials not cleared for reuse, all content of this type can be stored and distributed digitally to students with appropriate client display terminals and software. Assuming the availability of sufficient computer networking capability, these materials can be retrieved from storage as used by students with such terminals.

The storage requirements and level of technical expertise for electronic distribution of course materials are quite modest when the audio and video are not considered. However, strategies are required to ensure organization of materials in a useful way. A good example of the distribution of course materials of this type is demonstrated by Fox [11].

To support video storage, a complete 13 week course with three hours of instruction per week can be digitized and stored in less than 2.5 Gbytes at ISDN videoconferencing data rates (128 Kb/s). At MPEG-I data rates (1.5 Mb/s) this amount of video requires 26 Gbytes. However, a significant percentage of the video can be edited out as redundant (see Section 4). The current cost of disk storage (magnetic or read/write optical) is approximately $0.40/Mbyte, yielding video storage cost of less than $1,000 per course at ISDN data rates. Of equal importance is the ability of the server to meet the aggregate I/O bandwidth requirements of asynchronous student access [15]. In addition, access to stored material is likely to be uneven as the more difficult portions of a course will be accessed more often. This necessitates a multitiered memory hierarchy to reduce implementation costs [23].

3 Related Tools

Due to the multidisciplinary nature of educational and multimedia technology there is a great deal of existing work related to our requirements. Here we focus on recent tools and trials including the WWW, MBone, and distance learning systems that apply digital video.

3.1 The WWW in Distance Learning

A number or projects have been undertaken to put course materials on-line for student access using WWW technology. These include efforts at VPI&SU [11], Cornell [6], Global Network Academy [12], ongoing Digital Library Projects [9] and many more. All of these efforts, with some exceptions focus on distribution of materials and student content creation.

In the Multimedia Communications Laboratory (MCL) at Boston University we also have experimented with the use of the WWW to distribute course materials and to support student collaboration and presentations. Fig. 7 shows the "home page" used by students to access a variety of course materials including a syllabus, course reading, and student-contributed materials.

In this course we also used the WWW as a live teaching tool. This was achieved by a workstation equipped with a display scan converter to generate an NTSC video signal that was broadcast as conventional video. This enabled the capture and mixing of the computer screen with camera feeds in the classroom. This combination supported the creation of content for deferred access as well as the broadcast of lecture materials (Fig. 8).

Other related applications include the STREAMS project at Bellcore and the Stanford distance learning project. In the STREAMS prototype, users can record and playback technical presentations, seminars, and meetings [7]. During playback, STREAMS provides the user with an interface to visualize the session and locate specific events. The Stanford project addresses the design of a distance learning system to support a large student population [26]. The main focus of this effort is the design of the proper storage and network subsystems to

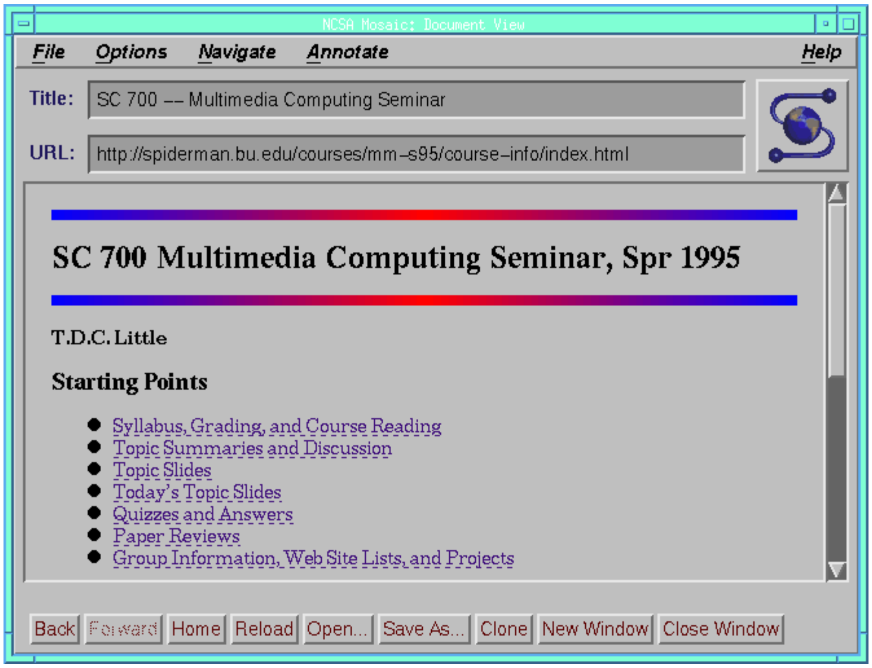

Figure 7: Example of WWW-Based Course Materials

Figure 8: Live Video Broadcast of Content Stored Using the WWW

support this endeavor.

3.2 The MBone

The use of the MBone for distance education has not been widespread primarily due to the lack of guaranteed bandwidth that results in poor audio and video quality. This problem is mitigated on isolated networks where bandwidth is freely available and is expected to change soon for public networks with the availability of network bandwidth reservation protocols [29]. There have been some trials on the use of the MBone for education. These include trials at the Naval Postgraduate School [17] and the Berkeley Multimedia Seminars [22]. Several conferences have been broadcast on the MBone, and the Internet Engineering Task Force periodically broadcasts all its workshops over the MBone.

4 Saddle

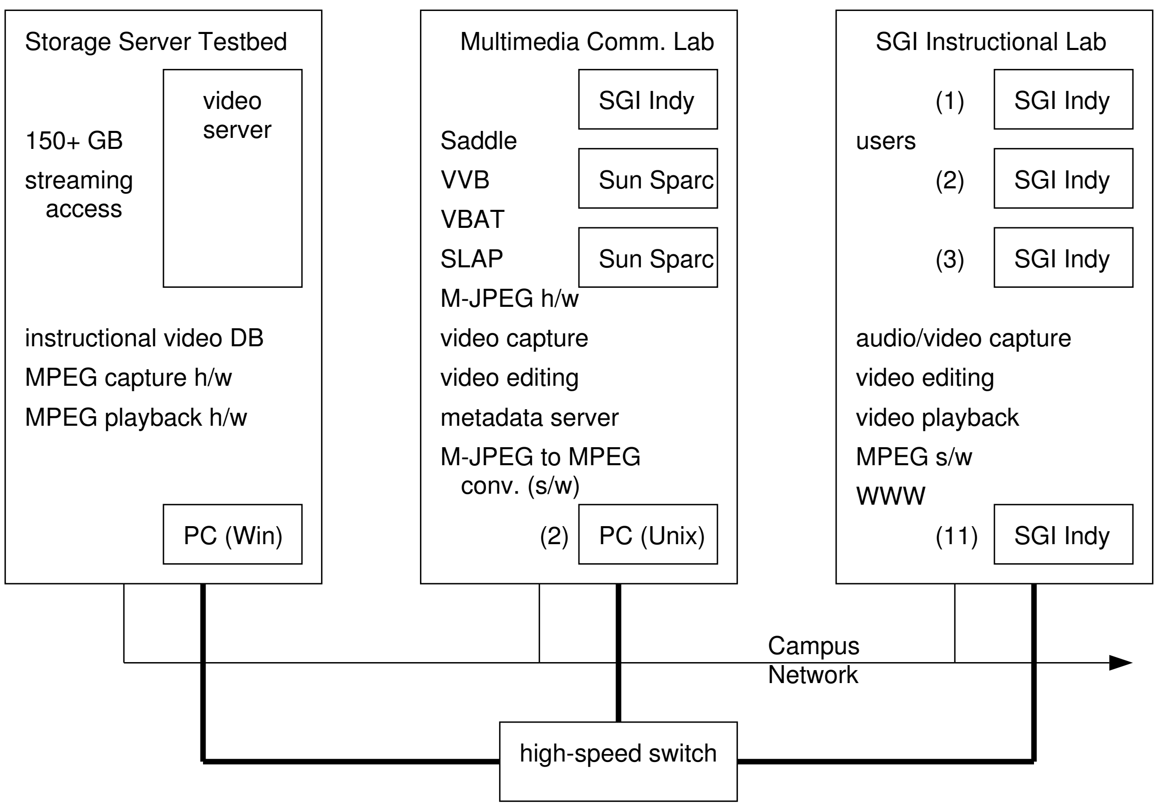

Figure 9: Proposed Saddle Testbed

Saddle (a scalable architecture for a distributed digital library environment) is a unifying research initiative that has grown out of our work in WWW content creation and multimedia application development. Because of our ready access to hundreds of hours of instructional video, we have targeted an educational scenario for one of its realizations. We envision a *personalized* system that can deliver courses to a student's desk without the need for a centralized receiving station as is currently used in microwave video broadcast. A student would then be able to participate in class from the office or home, or from any city. This system would allow wide-area accessibility to both synchronous (live) and asynchronous (stored) lessons, and would provide obvious benefits to its employer.

The Saddle data set is modeled on the graduate program at Boston University that services the local high-tech community via conventional analog video distance learning. This program has routinely produced 10-15 courses per year, all of which are recorded to video tape. The tapes are saved for several years, or until the space they consume must be reclaimed. In a single year over 500 hours of video instruction can be produced.

In the Saddle initiative we assume that this data set (or a portion thereof) can be digitized and put on-line, and thus supporting cross-referencing of instructional materials, self-paced and time-deferred study, and other use. The main objective is then to provide fast and interesting access to the information represented by the data set. The emphasis is to apply evolving communications and computer technology towards distance learning. This is motivated by the potential to immediately impact an existing base of distributed students. However, the same technology can be applied for on-campus instruction. In particular, the stored digital course materials can be used for on-line instruction during lectures via video projection equipment, or for self-paced, deferred instruction delivered to students on an individual basis using local or remote terminals.

In the remainder of this section we describe the construction of the Saddle system to meet the requirements outlined in Section 2.

4.1 The Saddle Testbed

The proposed testbed (Fig. 9) consists of content capture, authoring, storage, and playback functionality distributed over three components: (1) a large-scale storage server, (2) the MCL, and (3) the Software Engineering Lab. Here we leverage the use of new and existing equipment to provide playback stations and user evaluation.

The architectural design of the Saddle testbed is motivated by a number of technical evolutions including the popularity of the WWW as a means for information access and display, the availability of applications that allow video and audio to be played out in real time on workstations, and the multicasting capabilities provided by the MBone that support live audio, video and whiteboards.

The Saddle testbed is designed to be distributed in nature for flexibility and ease of access. The testbed functionalities are split among several components that are interconnected via a high-speed network. We are also motivated by the need to provide this learning infrastructure at a low-cost to its users. Our design relies on the use of existing components as much as possible to quickly develop the system. The design also emphasizes modularity to easily accommodate any new hardware/software into the system.

We now illustrate a hypothetical scenario in which a student can use the Saddle environment to initiate a learning session. (Acronyms are defined in Section 4.2.)

Saddle Operational Scenario: When a student initiates a session, the Saddle interface is displayed. The student can then decide to initiate a new query using the WWW/VVB and MovEase interfaces or recall a previous session using NavAid. The student also has the option of invoking the query interface at any point in the session. The browsing acts on the metadata and returns the results back to the student. If the student finds the proper information, a viewer can be invoked. The data are then identified from the metadata along with their location and are passed onto the viewer. The viewer then evaluates the availability of network and storage resources, establishes a connection to the database, and begins to deliver data to the student. This model of session control and delivery can be used for both live and stored media.

If the session is live, the user-initiated session is immediately aggregated into the proper multicast group to conserve I/O bandwidth. With stored media, sessions can be scaled to accommodate multiple data formats and batched to conserve resources [27]. In this scenario, both self-paced study and collaborative multicast modes are supported.

4.2 Current Efforts/Prototypes

The deployment of Saddle is aided by the development of several prototypes at the Multimedia Communication Laboratory and by a host of available shared research software. We now outline some of these efforts.

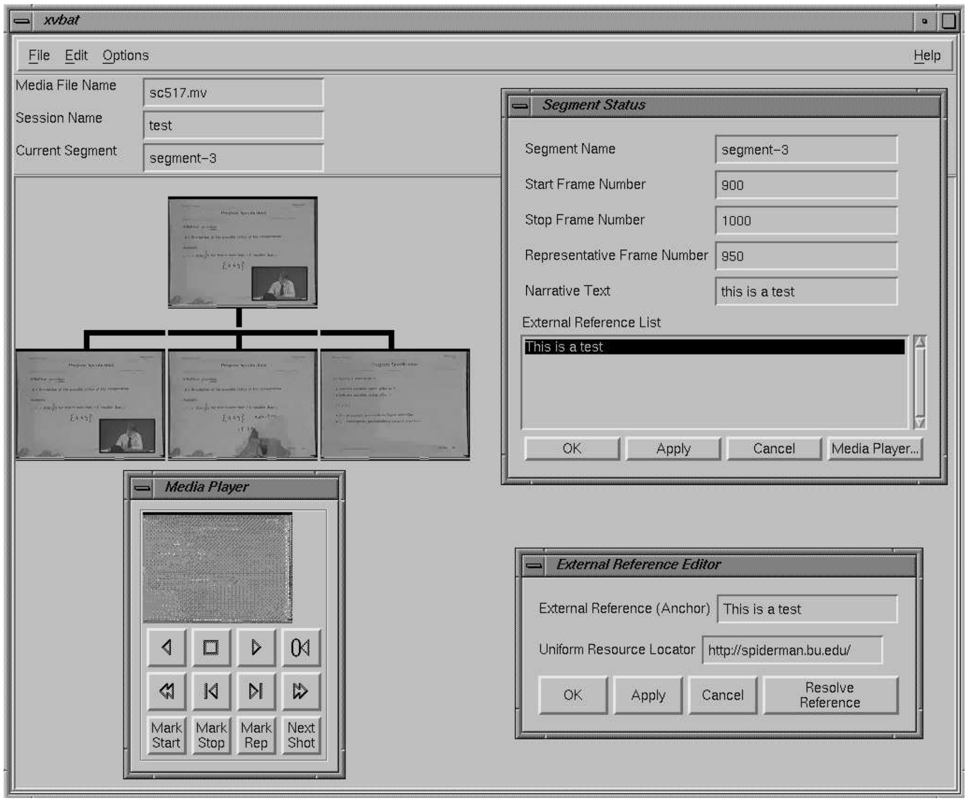

Figure 10: The VBAT Interface

The VBAT Application: The Video Broadcast Authoring Tool, or VBAT [4] inputs a set of parsed video segments, allows the user to dynamically rearrange and annotate indices in a hierarchical fashion, and outputs the results of the session in a segment description file (SDF). In addition, if applicable to a given browsing/playback environment, VBAT can create segmented pieces of the original video file, still images representing the indices, and other support files. This SDF taken as input to a given authoring session need not be the one provided by the acquisition system; VBAT supports multiple SDF's and sessions for a single media file, giving the user the capability to save a work in progress or create different views for different end users or playout systems.

VBAT's authoring paradigm is graphical and tree-based; users simply "drag-and-drop" objects representing segments of the media file into place on the screen (Fig. 10). Creation and deletion of segments is fully supported, and enhancements to the individual segments in the form of additional materials is provided through the specification of WWW Uniform

Resource Locators (URLs). In addition to URLs, per-segment data includes things such as start and stop frames, a representative frame to display as the visual identifier of a segment, a unique name, playout order, etc. VBAT also provides the user with authoring support services, including interactive designation of frames through a VCR-like playback facility, still image format translation, and automatic scene transition identification.

Scene detection in VBAT uses an approach similar to the one used in the Motion Picture Parser (MPP) application developed at the MCL [10]. For the existing MPP prototype we made several simplifications to the indexing and capture problem. We use a semi-automated approach to video indexing and content assignment. First, all text-based application-specific data are collected and organized for a video in a fixed database schema. The video is then parsed into components which are then assigned keywords and descriptive phrases to facilitate static content-based retrieval.

The VVB/WWW Instructional Database: The Virtual Video Browser (VVB) is an interactive VOD prototype designed to allow the browsing of a database of movies and the subsequent playout of individual movies [13]. It incorporates a simple query interface which lets users specify their preferences to the system to retrieve the appropriate video. The VVB is designed to work in a distributed environment where movies are stored on multiple distributed servers interconnected via a network. Its behavior is typical of distributed multimedia applications. The VVB employs a metadata location mechanism. The metadata mechanism and supporting protocols are based on the client-server model. The current WWW prototype interface for Saddle, based on the VVB/WWW prototype is shown in Fig. 11. In this version, small stills act as buttons to access video clips corresponding to individual topics.

MovEase: To perform efficient queries on a database of multimedia information, efficient mechanisms for retrieval are required. A flexible and intuitive interface for query formulation is required along with a good indexing structure. We seek to integrate image, audio, and text processing tools into a generic information filter to extract content. MovEase is a tool used to prototype some of our initial work in motion query formulation [1].

The MovEase interface is divided into three regions: work area or query builder, icon catalog browser, and the result browser. A query is composed in the query builder, the contents of the object, motion, or query catalog are viewed in the icon catalog browser, and the results of a query (the data retrieved from the database) can be evaluated in the result

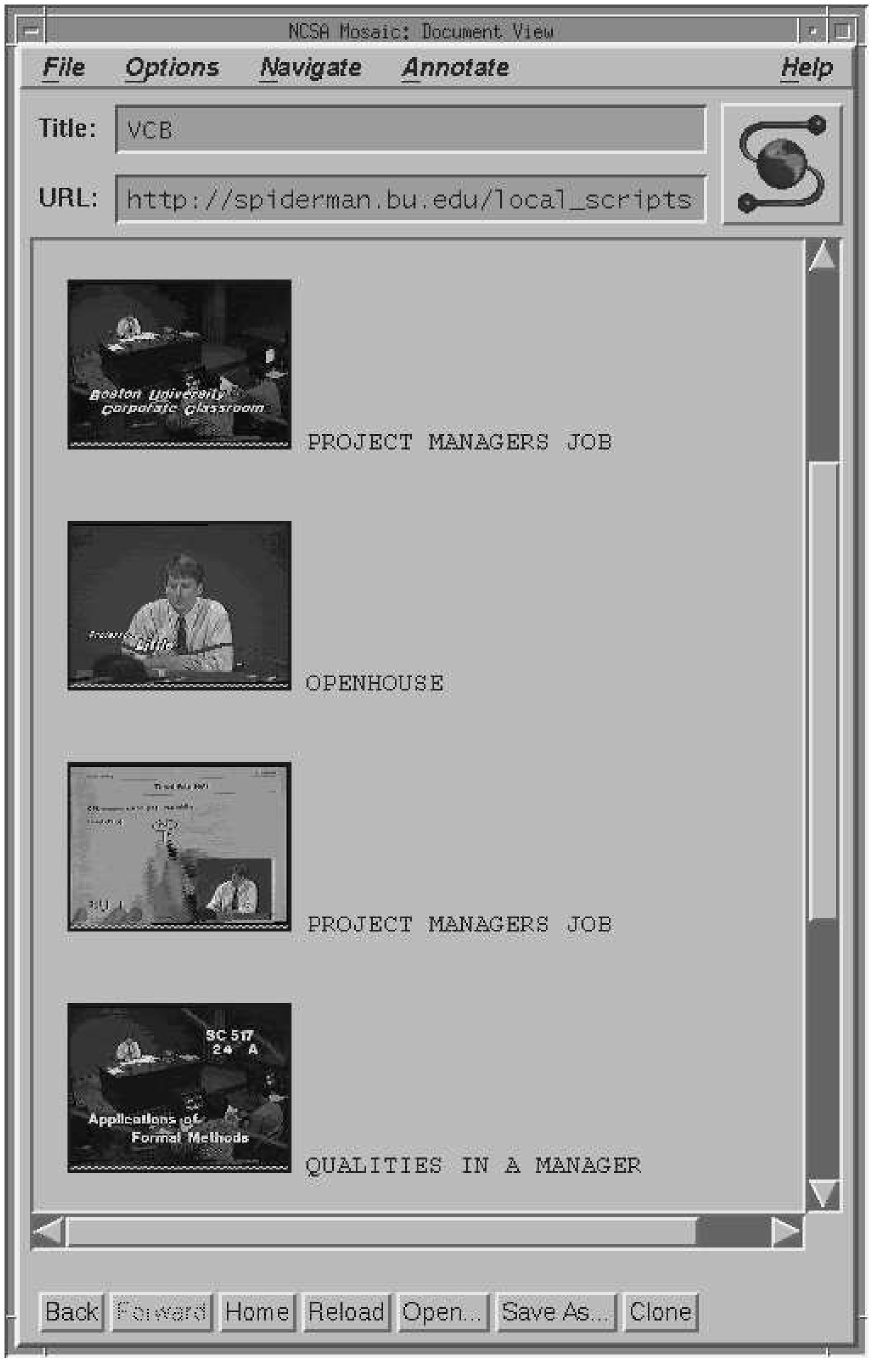

Figure 11: A Distance Learning Interface

browser.

CHAMP: To alleviate the deficiency of batch-mode transfers provided by HTTP, we have developed a software-only video client/server pair called CHAMP (Continuous Hyper-Application Media Player). This software component is used for the delivery and playout of M-JPEG-compressed video in conjunction with the VVB. The player provides a full-function VCR graphic user interface for simple manipulation of playout. The application is optimized for stored video and is therefore hardware independent at playout time. Playout at the client side can be performed either through software decompression or using specialized decompression hardware. Flow-control and congestion control are achieved through a combination of frame-dropping and quality-scaling as required by the application.

NavAid: NavAid is a tool that lets users navigate documents interconnected via WWW hyperlinks. NavAid provides a graphical interface that lets the user visualize the relationships among interconnected documents. The documents appear as a series of nodes in a graph as the user navigates the database. The user has the option of returning to a document by simply clicking on the appropriate node. Such an interface can be very useful in navigating a complex database.

Other Tools: To reach the widest audience, the Saddle prototypes are being designed to interface with existing Internet software including the WWW and the MBone. The capture and index tools in Saddle are being designed to work with MBone data formats and interfaces (e.g., nv, vat, vic, and wb) for wide accessibility.

4.3 Research Issues

We now discuss briefly some additional research problems that are being addressed in the construction of the Saddle prototype.

Data parsing to segments (clips) for easy access and retrieval: Although a great deal of information can be contained in audio and video data (e.g., newscasts), the information content is difficult to characterize and to extract [24, 25]. Both static and dynamic approaches have been developed to respond to queries about the content of a data segment (e.g., "Find all newsclips on health care" or "find the shots with homework solutions").

To achieve this goal we must recognize the relationships among the capture methods, the application models, the database schemata, the retrieval engine, and the data interchange formats [2, 3, 8]. To this end, requirements in this domain are:

1. Identification of application-specific information models for educational materials, instructional videos, and seminars.
2. Integration of feature and content extraction methods to support both batch-mode and dynamic processing modes.
3. Development of tools to facilitate the rapid input and indexing of composite audio, video, and text-based information content.
4. Development of tools to facilitate authoring of multiple media elements and information content (bypassing extraction phase).

With respect to image and video databases, two schemes appear to be evolving to achieve the content-based retrieval component. The first approach requires preprocessing the contents of the database to glean domain-specific information into readily-machine-recognizable tokens (text) for indexing and subsequent query. However, this approach cannot anticipate all possible types of queries that might be applied to the database. The second approach is dynamic. Queries are formulated in mixed formats and applied to database as a whole rather than to the pre-extracted indices. For audio, image, and video databases, both approaches require significant use of image/signal processing. For the latter approach, flexibility is traded-off for performance. A more reasonable solution would provide a balance between a priori content-skimming and dynamic content identification from the objects in the database.

We are interested in both on-line techniques and off-line processing methods for live and post-production processing of video clips. Our approach focuses on the use of snippets of video (vicons) and image miniatures (picons) to represent data for processing which simplifies the rendering process. On-line techniques include automatic indexing (which can be time or content based) and provision for the placement of markers by the author (instructor) to help in the post-processing and editing phase. This approach is similar to that used in the Continuous Media Toolkit [21].

A related issue is the use of the structure of the instructional environment to improve the process on on-line indexing and tagging techniques. Videos recorded in the classroom have a regular structure and format. Typical camera views include:

- The instructor's notepad,
- a head-and-shoulders of the instructor,
- a panoramic view of the class
- a combination of the head-and-shoulders iconified and superimposed onto the instructor's notepad, and
- daily leaders and trailers for identifying the tapes.

Such a structured representation of video allows easy segmentation as has been demonstrated by Zhang et al. [28] but is useless without contextual information. We are experimenting with automatic tagging techniques with which the instructor can mark the video while it is being recorded to simplify the post-processing. For example, the instructor can use different colored pads to introduce new topics or electronically tag an important or new topic in the video stream. In the future, more sophisticated techniques such as image content and voice recognition techniques can be used to automate the indexing process [18, 20].

A related issue is the evolution from video of hardcopy to on-line materials as visual aids for ease of access. This includes iconic views of the class, instructor, and the notepad to help the student select the appropriate topic.

Representation of Timing: Efficient processing can reduce the system requirements but can lead to synchronization problems. Timing represents a core component of the target application as video and audio data are time-dependent. It is critical that the presentation of the instructional material be synchronized. Some of the issues that can lead to synchronization problems include:

- deletion of redundant audio (e.g., the instructor uses the same audio segment in different part of the lecture),
- deletion of redundant video (compression),
- replacement of video with representative stills, and
- substitution of text (e.g., PostScript images) for video.

When we replace video or audio there is a potential to lose timing information that must be preserved. For example, the replacement of 60 seconds of video by a PostScript-encoded still yields significant compression but introduces a timing problem: the still must be associated with 60 seconds of the video soundtrack. We are integrating our existing timing representations for this purpose [16].

Data Storage and Serving: This problem includes researching into the various data layout and placement strategies. Specifically, we are investigating the memory hierarchies and data placement policies necessary to satisfy the requirements of a large student population. We also wish to satisfy the needs of the instructor and find a mapping between this creation of content and retrieval.

Data Delivery, Scaling and Synchronization: In this component, we are investigating the various network protocols necessary to support the delivery of information to students at remote locations. We are also studying the use of scalable video compression techniques to satisfy the playout requirements of a diverse set of clients. Other issues include the use of multicast protocols for multipoint delivery, pricing policies and session synchronization.

Service Group Scaling We are also investigating issues related to the dynamic batching of multicast sessions to conserve network and server I/O bandwidth. Consider the scenario in which an instructor begins a lecture at 1:00PM. Due to the limited server I/O bandwidth and network connection service, access to course and course database is restricted to a window of opportunity during peak hours (e.g., mid-day). As the course is recorded on the fly, students can join a *service group* for the live multicast or can join late, forming a new group skewed by 10 min. Students who join late can "catch up" by starting interactive point-to-point sessions and browsing the recorded content and moving through it at a quicker pace. Ultimately they *demote* their interactive sessions by joining the live service group. As other students arrive late to class, they form similar groups, perhaps choosing to remain on their skewed tracks. When the instructor presents a reference to the previous class, some students may decide to access the video recorded from the previous class. This group can be either aggregated into a single session or spawned out as individual point-to-point connections, depending upon the availability of network resources. Near (in time) sessions are demoted by aggregation into service groups or a subset of the point-to-point sessions are provided with substandard service. We are currently developing protocols to achieve this functionality [27].

5 Summary

Distance learning technology promises to enable widespread access to instructional content provided from digital libraries and institutions of higher education. In this paper we have described the requirements for the construction of a distance learning system supporting large-scale digital video storage and distribution. We also have presented the architecture of our evolving distance learning testbed in the Multimedia Communications Lab. To date this unifying effort has provided a rich set of research problems related to multimedia communications and promises to bring new insight to computer-based education as it evolves to accommodate lessons learned in its application.

References

[1] G. Ahanger, D. Benson, and T.D.C Little, "Video Query Formulation," *IS&T/SPIE Symposium on Electronic Imaging Science and Technology (Storage and Retrieval for Image and Video Databases III)*, San Jose, CA, February 1995, SPIE Vol. 2420, pp. 280-291.

[2] T.G. Aguierre Smith and G. Davenport, "The Stratification System: A Design Environment for Random Access Video," *Proc. 3rd Intl. Workshop on Network and Operating System Support for Digital Audio and Video*, La Jolla, CA, November 1992.

[3] F. Arman, A. Hsu, and M-.Y. Chiu, "Image Processing on Compressed Data for Large Video Databases," *Proc. 1st ACM Intl. Conf. on Multimedia (ACM Multimedia'93)*, Anaheim CA, August 1993, pp. 267-272.

[4] M. Carreira, J. Casebolt, G. Desrosiers, and T.D.C. Little, "Capture-Time Indexing Paradigm, Authoring Tool, and Browsing Environment for Digital Broadcast Video," *Proc. IS&T/SPIE Symposium on Electronic Imaging Science & Technology*, San Jose, CA, February 1995.

[5] S. Casner, "Are You on the MBone?" *IEEE Multimedia*, Vol 1., No. 2, Summer 1994, pp. 76-79.

[6] Cornell Theory Center, "Education," *http://www.tc.cornell.edu:80/Edu/*.

[7] G. Cruz and R Hill, "Capturing and Playing Multimedia Events with STREAMS," *Proc. 2nd ACM Intl. Conf. on Multimedia, (ACM Multimedia'94)*, San Francisco, CA, October 1994, pp. 193-200.

[8] G. Davenport, T.G. Aguierre Smith, and N. Pincever, "Cinematic Primitives for Multimedia," *IEEE Computer Graphics & Applications*, July 1991, pp. 67-74.

[9] The Digital Library Initiative, *http://www-diglib.stanford.edu/diglib/pub/*.

[10] E. Deardorff, T.D.C. Little, J.D. Marshall, D. Venkatesh, and R. Walzer, "Video Scene Decomposition with the Motion Picture Parser," *IS&T/SPIE Symposium on Electronic Imagery Science & Technology (Digital Video Compression and Processing on Personal Computers: Algorithms and Technologies)*, San Jose, February 1994, SPIE Vol. 2187, pp. 44-55.

[11] E.A. Fox, "Interactive Learning with a Digital Library in Computer Science," *http://fox.cs.vt.edu/ EI.html*, Dept. of Computer Science, Virginia Polytechnic Institute and State University.

[12] Global Network Academy, "Introduction to Object Oriented Programming Using C++," *http://uu-gna.mit.edu:8001/uu-gna/text/cc/ index.html*, 1994.

[13] T.D.C. Little, G. Ahanger, H.-J. Chen, R.J. Folz, J.F. Gibbon, A. Krishnamurthy, P. Lumba, M. Ramanathan, and D. Venkatesh, "Selection and Dissemination of Digital Video via the Virtual Video Browser," *Journal of Multimedia Tools and Applications*, Kluwer Publications, Vol. 1, No. 2, June 1995, pp. 149-172.

[14] T.D.C. Little and D. Venkatesh, "Prospects for Interactive Video-on-Demand," *IEEE Multimedia*, Vol. 1, No. 3, Fall 1994, pp. 14-24.

[15] T.D.C. Little and D. Venkatesh, "Probabilistic Assignment of Movies to Storage Devices in a Video-on-Demand System," *ACM/Springer Multimedia Systems*, Vol. 2, No. 6, January 1995, pp. 280-287.

[16] T.D.C. Little and A. Ghafoor, "Interval-Based Temporal Models for Time-Dependent Multimedia Data," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 5, No. 4, August 1993, pp.551-563.

[17] M.R. Macedonia and D.P. Brutzman, "MBone Provides Audio and Video Across the Internet," *Computer*, Vol. 27, No. 4, April 1994, pp. 30-36.

[18] J. Matthews, P. Gloor, and F. Makedon, "VideoScheme: A Programmable Video Editing System for Automation and Media Recognition," *Proc. 1st ACM Intl. Conf. on Multimedia (ACM Multimedia'93)*, Anaheim CA, August 1993, pp. 419-426.

[19] J.S. Noblitt, "Enhancing Instruction with Multimedia," *Proc. 4th Natl. Conf. on College Teaching and Learning*, Center for the Advancement of Teaching and Learning at Florida.

[20] K. Otsuji and Y. Tonomura, "Projection Detecting Filter for Video Cut Detection," *Proc. 1st ACM Intl. Conf. on Multimedia (ACM Multimedia'93)*, Anaheim CA, August 1993, pp. 251-257.

[21] L.A. Rowe et al., "The Continuous Media Toolkit," *http://www-plateau.cs.berkeley.edu/cmt /index.html*, University of California, Berkeley, 1995.

[22] L.A. Rowe and M. Vetterli, "CS 298-5: Multimedia and Graphics Seminar," *http://roger-rabbit.cs.berkeley.edu/298.html,* Dept. of EECS, University of California, Berkeley.

[23] L. Rowe, J. Boreczky, and C. Eads, "Indexes for User Access to Large Video Databases," IS&T/SPIE Symposium on Electronic Imagery Science & Technology (*Digital Video Compression and Processing on Personal Computers: Algorithms and Technologies)*, San Jose, February 1994, Vol. 2187, pp. 1-12.

[24] D. Swanberg, C.F. Shu, and R. Jain, "Architecture of a Multimedia Information System For Content-Based Retrieval," *Proc. 3rd Intl. Workshop on Network and Operating System Support for Digital Audio and Video*, La Jolla, CA, November 1992.

[25] D. Swanberg, T. Weymouth, and R. Jain, "Domain Information Model: An Extended Data Model for Insertions and Query," *Proc. 1992 Workshop on Multimedia Information Systems*, Tempe, Arizona, February, 1992, pp. 39-51.

[26] F. A. Tobagi, "Distance Learning with Digital Video," *IEEE Multimedia*, Vol. 2, No. 1, Spring 1995, pp. 90-93.

[27] D. Venkatesh and T.D.C. Little, "Dynamic Service Aggregation for Efficient Use of Resources in Interactive Video Delivery," *Proc. 5th Intl. Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV'95)*, Durham, New Hampshire, April 19-22, 1995, pp. 119-122.

[28] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic Partitioning of Full-Motion Video," *Multimedia Systems*, Vol. 1, No. 1, 1993, pp. 10-28.

[29] L. Zhang, R. Braden, D. Estrin, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification," *Internet Draft draft-ietf-rsvp-spec-03.ps,* Xerox/ISI/USC, July, 1994.