Media Streams: Representing Video for Retrieval and Repurposing (Demonstration Description)

Bibliographic Reference:

Marc Davis. "Media Streams: Representing Video for Retrieval and Repurposing (Demonstration Description)" In: *Proceedings of Second ACM International Conference on Multimedia in San Francisco, California*, ACM Press, 464-465, 1994.

ACM Multimedia '94 Demonstrations Program

T.D.C. Little
Boston University*

Overview

The demonstration program features novel research prototypes that demonstrate the latest advancements in multimedia computing and communications technologies. We solicited proposals for demonstration of working systems that highlight the essence of multimedia: the integration of technologies and media. Each demonstration proposal was reviewed by at least three referees. Twelve proposals were selected out of thirty submissions. Equipment requirements were not a factor in the final acceptances. In general, the referees favored (1) new and demonstratable concepts or prototypes, (2) interesting "content" applied using multimedia technology, and (3) interesting and new integrations of existing technologies. A number of the proposals were forwarded to the video and product exhibits programs.

The theme of the demonstrations program is information indexing, browsing and delivery, and focuses on visual applications. The demonstrations will be exhibited at regular intervals by their creators with time set aside for personal interaction with the systems. Accepted demonstrations fall, roughly, into the following categories:

- Video delivery
- Information indexing and browsing
- Authoring tools

Format of Demonstrations

Wednesday morning (October 19) will be used for set-up in a single demonstrations room. The program will run Wednesday noon until 6.30 PM and Thursday 9:00 AM until 12:30 PM. Thursday afternoon is reserved for equipment tear-down.

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

Demonstrations will be run in two modes: (1) formal, scheduled group presentations lasting approximately 15-20 minutes, and (2) informal, interactive and unscheduled presentations. Four concurrent presentations will be scheduled on the hour, 20 minutes past the hour, and 40 minutes past the hour. (One formal presentation per demonstration will be scheduled per hour.) The scheduled times of each presentation will be posted and advertised.

Acknowledgements

I thank the Demonstrations Program Committee members for their timely reviews of the demonstration proposals. Special thanks go to Jess Gibbon who provided video tape duplication.

The Demonstrations

The demonstrations are summarized below.

1. Multimedia Memory of Mayan Medicine
2. The ViewStation Project: Computer-Participative Media Applications
3. Media Streams: Video Representation for Retrieval and Repurposing
4. Galaxy of News and Information Landscapes: Dynamic Visualization and Access of Information in a Multidimensional Space
5. MUDE: A Multimedia Database Engine for Content-Based Retrieval
6. Capturing and Playing Multimedia Events with STREAMS
7. Informedia Digital Video Library
8. The Virtual Video Browser in Mosaic
9. On-Line Distance Learning
10. Multimedia Teleconferencing on the Internet Multicast Backbone (MBONE)
11. Distributed Music Trio – Media Synchronization Over Wide Area
12. CMIFed: A Transportable Hypermedia Authoring System

The individual demonstrations are described in the remainder of this document.

1 Multimedia Memory of Mayan Medicine

Thomas Aguierre Smith
MIT Media Laboratory
Barbara Woloch
School of Film and Television, NYU
Laura Teodosio
MIT Media Laboratory
Augustine Farias C., Juan Carlos de la Parra, and Feliciano Gómez Santíiz (Production Associates)
Centro de Investigaciones Ecológicas del Sureste de Mexico

Contact: *teo@media.mit.edu*

Synopsis: Ethnobotanical data incorporates many different facets of knowledge and experience. Much of the classification and nomenclature is associated with both visual and aural aspects of the environment. Three years ago, a program for the recording and dissemination of Tzeltal and Tzotzil Maya ethnobotanical and ethnomedical knowledge was initiated by The Maya Audio Visual Project in association with PROCOMITH (Programa de Colaboracion Medicina Indigena Tradicional Herbolaria) and C.I.E.S. (Centro de Investigaciones Ecológicas del Sureste de Mexico). Our demonstration will present the results of audio/visual fieldwork conducted in Highland Maya communities.

The focal point for all the various types of information that will be presented are the concepts of "place" and "narrative." "Place" is inclusive of all natural phenomena, meaning both the material environment as well as the history of the given area. The term "narrative" is used to define social interactions among

Maya healers and their patients as well as interactions between people and plants. While the relationship between "place" and "narrative" is a circular one, each influencing the other in turn, the two are separated for the purposes of study. In our configuration of the world of Mayan healing, "place" is represented by a quadrilateral which encompasses healers, patients, illnesses and plants. Each quadrilateral functions as a discursive space for our presentation of audio visual materials from a given area. "Narratives" link the various elements. The elements include case studies of illnesses, as well as the interactions between healer and patient and person and plant. The narratives allow for continuity between the four nodes of information as well as an understanding of how these four elements are interrelated.

We have developed a multimedia, interactive, computer-based information system to describe and evaluate therapeutic practices of Maya herbal specialists in Chiapas, Mexico. We employ advanced computer technology combined with sensitive anthropological field work to (1) document ethnopharmaceutical preparation of Maya herbal remedies with demonstrated pharmacological activity and (2) to develop and test hypotheses concerning possible chemical constituents relating to pharmacological efficacy of Maya herbal prescriptions.

2 The ViewStation Project: Computer-Participative Media Applications

David Tennenhouse, Joel Adam, Chris Lindblad, William Stasior, and David Wetherall
Telemedia Networks and Systems Group, Laboratory for Computer Science, MIT

Contact: *djw@lcs.mit.edu*

Synopsis: The ViewStation is a local-area distributed multimedia system that uses a high-speed ATM network to combine multimedia peripherals and Unix workstations in an object-oriented application environment. A key tenet of the project is the delivery of media data not just to the desktop but all the way to the application program. In ViewStation applications, the computer is an active participant and performs intelligent processing that depends on the content of the video.

We have found that the ViewStation provides a good platform for investigating concrete ways that computers may become more responsive to their human users. We demonstrate components of a prototype "Computerized Office Multimedia Assistant" (COMMA), which assists its user by performing various tasks that require the analysis of live video. The Room Monitor component uses video clips to summarize activity in a room that is monitored over a period of time, while the Whiteboard Recorder keeps a history of changes to an office whiteboard.

Other ViewStation applications digest produced video and demonstrate the feasibility of agents that perform content-based media processing. The Sports Highlight Browser uses graphical templates to segment a recorded sporting news telecast into a set of video clips, each of which represents highlights of a particular sporting event. The Joke Browser uses closed captions associated with the video of late-night talk show monologues to segment them into individual jokes that may be browsed by topic. The News Browser provides interactive access to a simple database of broadcast television news articles, compiled using assumptions about the format of the broadcasts.

3 Media Streams: Representing Video for Retrieval and Repurposing

Marc Davis
MIT Media Laboratory and Interval Research Corporation

Contact: *davis@interval.com*

Synopsis: In order to enable the search and retrieval of video from large archives, we need a representation of video content. Although some aspects of video can be automatically parsed, a representation sufficient for content-based retrieval requires that video be annotated. Over the past three years, members of the MIT Media Laboratory's Machine Understanding Group in the Learning and Common Sense Section (Marc Davis with the assistance of Brian Williams and Golan Levin under the direction of Prof. Kenneth Haase) have been building a prototype for the annotation and retrieval of video data. This system is called Media Streams.

Media Streams is written in Macintosh Common Lisp and FRAMER, a persistent framework for media annotation and description that supports cross-platform knowledge representation and database functionality. Media Streams runs on an Apple Macintosh Quadra 950 with two high resolution, accelerated 24-bit color displays and uses Apple's QuickTime digital video format. With Media Streams users create stream-based, temporally-indexed, iconic annotations of video content which enable content-based retrieval of annotated video sequences.

The Icon Space is the interface for the selection and compounding of the iconic descriptors in Media Streams. To date there are over 3000 iconic primitives. Through compounding, the base set of primitives can produce millions of unique expressions. In the Icon Space, users can create palettes of iconic descriptors for use in annotation and search. By querying the space of descriptors, users can dynamically group related iconic descriptors on-the-fly. These icon palettes enable users to reuse the descriptive effort of others.

The Media Time Line is the core browser and viewer of Media Streams. It enables users to visualize video at multiple timescales simultaneously, to read and write multi-layered iconic annotations, and provides one consistent interface for annotation, browsing, query, and editing of video and audio data. Queries can be formulated either by description or by example. In query by description, one simply uses the Media Time Line to describe the desired video shot or sequence. Annotation is describing video one has; query is describing video one wants to find. In query by example, on uses an existing annotated video shot or sequence as a query itself and retrieves video shots or sequences similar to the example.

With Media Streams users can create shareable representations of media content which enable the construction of large archives of reusable temporal media. Without tools like Media Streams, a thousand hours of video content will be less useful than one. With tools like Media Streams, all the digital video in the world can become an accessible and reusable resource for the human and computational imagination.

4 Galaxy of News and Information Landscapes: Dynamic Visualization and Access of Information in a Multidimensional Space

Earl Rennison
Visible Language Workshop, MIT Media Laboratory

Contact: *rennison@media.mit.edu*

Synopsis: The Galaxy of News system embodies a scalable approach to visualizing and navigating through large quantities of independently authored pieces of information, in this case news stories. It combines the effective aspects of both searching and browsing and the ability to switch between these modes of operation seamlessly within a single interface. The system automatically organize disconnected articles into dynamically formed groups, based on the content of the articles, that allow quick access to related information and the ability to quickly understand the relationships between articles.

Galaxy of News creates a new medium, an abstract information space, providing interactive navigation and intuitive access to correlated information. The Galaxy of News information spaces form structured, multidimensional, interactive environments where the information objects contained within the space determine the underlying structure. A powerful relationship construction engine utilizes an associative relation network to automatically build implicit links between related articles. To enhance the understanding of the space and its contents, the multidimensional information spaces constructed can change scale, orientation, perspective, representation and presentation as the user navigates through the space. Users interact with these information spaces through visual dialogs where actions have implicit meaning, e.g. moving forward in the space indicates more specific detail is desired and moving backward indicates a desire for less detail and more abstraction.

The Galaxy of New project investigates several information access and visualization principles, including (1) pyramidal visualization of news objects to provide progressive refinement of news information, (2) visual clustering of news elements based on the content of news articles to provide structured information access, (3) semantic zooming and panning, where zooming is synonymous with searching or filtering, and panning is synonymous with browsing, (4) fluidity of interaction to understand and maintain the context of the information being presented, (5) animation and motion to illustrate relationships between news elements, (6) dynamic visual cues to aid in the navigation through an abstract news information space, and (7) dynamic visual presentation of information to present the proper quantity of information at each instance of interaction and to eliminate distracting clutter. These principles define an outline for building a structured hierarchical representation of news, whereby the upper portions of the pyramid consist of general descriptions or abstractions of the lower levels which contain more detail. Pyramidal representation offers news readers the ability to go through a process of glancing, to investigating, to reading details in a fluid and selective manner, while maintaining context of where they are in the process.

The Galaxy of News project is part of a larger research effort in the Visible Language Workshop exploring Information Landscapes. This research explores new methods of navigating and browsing includes the use of transparency, blur, layering, and infinite zoom. New technologies allow us to swoop in and out, rotate around the information – to browse it, to drill in for depth and detail as well as to grasp its organization. The ideas shown in these demonstrations – such as typographical constructs, transitional movements, information transforms, multiple points of view, and interconnected pathways – combine with our previous work in transparency, blur layering, adaptive graphics and infinite zoom. They become building blocks for a new language of visualization. In this world, traditional design principles must be re-examined and new vocabularies invented to be consonant with this new multidimensional information environment.

5 MUDE: A Multimedia Database Engine for Content-Based Retrieval

J. K. Wu, A. Desai Narasimhalu, Y. H. Ang, C. P. Lam, H. H. Loh, and B. Mehtre
Institute of Systems Science, National University of Singapore

Contact: *jiankang@iss.nus.sg*

Synopsis: With rapid advances in image database research and development, the need for a generic multimedia database engine which can readily be customized for various image database applications is increasing. We present a MUltimedia Database Engine (MUDE) in this demonstration It consists of three modules: image processing, indexing and retrieval, and storage manager for large systems. The database engine is based on an image representation hierarchy, which contains four levels: attribute, feature and description, segmented image, and image data. The image processing module provides various image processing and editing functions to derive a complete representation. Indexing and retrieval module provides several indexing techniques and multiple access methods to image and non-image data. An iconic indexing method is presented which uses a special neural network model to fuse multi-modal features and to spatially self-organize nodes of an indexing tree. The database engine MUDE has been developed and used for two applications: a Computer-Aided Facial Image Inference and Retrieval (CAFIIR) and a System for Trademark Archival and Registration (STAR).

CAFIIR system was developed for a mugshot application. Unique features of CAFIIR, namely, facial image composition, image normalization and feature extraction, multiple retrieval and indexing techniques, facial image aging, will be demonstrated. Besides traditional database indexing techniques, CAFIIR uses classification and clustering techniques for generating additional indexes. It supports retrieval of faces by regular database attributes such as the name and ID number. In addition, it also allows for retrieval of faces either through indexing tree browser, fuzzy descriptors or free text descriptors.

Trademarks are complex patterns consisting of various image and text patterns, called *device-mark* and *word-in-mark* respectively. Traditionally, only text part has been used for search and retrieval of such patterns. This was largely due to the diversity and complexity of image patterns occurring in trademarks. STAR uses multiple features based on image/device-mark (shape, structure, complexity, symmetry, meaning) as well as word-in-mark (text, phonetics, interpretation) and brings out the conflicting trademarks for the consideration of trademark officer. Unique features of the system, namely, structured segmentation, searching with composite similarity measures and thesaurus, etc. will be demonstrated.

6 Capturing and Playing Multimedia Events with STREAMS

Gil Cruz and Ralph Hill
Bell Communications Research

Contact: *gil@bellcore.com, rdh@thumper.bellcore.com*

Synopsis: STREAMS is a prototype application designed and implemented at Bellcore to support the recording and playback of technical presentations, training sessions, and meetings. During playback STREAMS lets users make choices about how the recorded information is to be presented. To further aid users, STREAMS incorporates powerful searching techniques for locating information in audio and video streams. STREAMS is based on digital storage of multiple streams. A stream is typically an audio track or a (silent) video track. Streams are stored independently, but share a common dimension - time. This common dimension is used to synchronize the streams during playback, browsing, and searching.

The driving paradigm of STREAMS is to empower the user to make intelligent choices. STREAMS lets the user (implicitly) allocate resources by controlling playback. For example, the user can turn off playback of a video stream, increase the quality of the audio, or decrease the frame size of another video stream. STREAMS also empowers the user by providing searching mechanisms designed to enable rapid browsing so users can quickly identify points of interest.

Attendees at the demo will be able to use the STREAMS system to browse various talks that have been captured at Bellcore. In addition we will attempt to capture the opening plenary panel at the conference and allow attendees to search and browse the STREAMS record of this panel.

7 Informedia Digital Video Library

Michael Christel, Scott Stevens, and Howard Wactlar
Software Engineering Institute, Carnegie Mellon University

Contact: *mac@sei.cmu.edu*

Synopsis: The Informedia Project is establishing an extremely large on-line digital video library, initially incorporating video assets from WQED/Pittsburgh. The project is creating intelligent, automatic mechanisms for populating the library and allowing for its full-content and knowledge-based search and segment retrieval. The library's contents are conveyed through both narrative (speech and language) and images, and hence it is segmented and indexed through the integration of speech, image, and natural language understanding technologies.

The demonstration shows the exploration of a digital video library consisting of material from WQED's "Space Age" series. It illustrates the utility of indexing video through text transcripts, intelligent searching with natural language, and video segmentation based on content. The demonstration shows how these concepts can result in faster and more efficient retrieval of relevant information, but is limited in that the data comes only from a local gigabyte

drive, and it shows a first generation user interface utilizing only keyboard query. Forthcoming improvements will be access to data transmitted by high speed networks, and empirically validated, age-appropriate user interfaces with multimodal query support. Whereas the creation of the "Space Age" digital video library required human support for scene segmentation and transcript generation, future library creation will be streamlined through improved integration of image and natural language processing for automated scene segmentation and speech recognition for automated transcript generation.

8 The Virtual Video Browser in Mosaic

Dinesh Venkatesh and T.D.C. Little
Multimedia Communications Laboratory, Boston University

Contact: *mcl@spiderman.bu.edu*

Synopsis: The Virtual Video Browser is a true-VOD (video-on-demand) application that permits an individual to browse an electronic video collection stored in distributed video database. The basic objective in the application development was to illustrate the capabilities of a temporal modeling scheme and to gain experience with an interesting multimedia application that requires the convergence of many technologies. Secondary objectives were to design a software system with various desirable properties including ease of use.

The VVB operates in association with a metadata server. Movies and scenes of movies can be identified and viewed through the VVB based on movie-specific attributes including actor names, director names, and scene characteristics. Movie and scene content indexing are facilitated by summary, keyword, and transcript searching. In addition to viewing scenes from the movies, a user can view all of a movie's textual information including summaries and transcripts. The VVB 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 in which movies are stored on different video servers interconnected via a network.

In the demonstration we present a version of our Video-on-Demand (VOD) application developed using the HTML syntax and running on the World Wide Web. The application was originally built to operate using a Unix workstation equipped with video compression and playback hardware. The VVB interface was re-implemented using the HTML forms interface and relies on the metadata server of the original VVB. Movie playout is facilitated by the Berkeley MPEG player and the conversion of our motion-JPEG movies to MPEG files. The VVB in Mosaic can be accessed via the Web at http://spiderman.bu.edu.

9 On-Line Distance Learning

Fouad A. Tobagi
Stanford University

Contact: *tobagi@sierra.stanford.edu*

Synopsis: This is a technical demonstration showing the integration of several technologies (real-time MPEG compression, video servers, digital video communication and networking, live and stored playback of digital video, and user interfaces) as applied to distance learning.

10 Multimedia Teleconferencing on the Internet Multicast Backbone (MBONE)

Stephen Casner and Eve Schooler
USC/Information Sciences Institute
Ron Frederick
Xerox Palo Alto Research Center
Van Jacobson
Lawrence Berkeley Laboratory

Contact: *casner@isi.edu*

Synopsis: Worldwide teleconferencing with 10's or 100's of participants is now possible over the Internet MBONE using audio, video and shared whiteboard tools available for free.

During the past two years, there has been dramatic growth in the development and use of real-time audio and video teleconferencing over the Internet. This growth began with the transmission of live audio and video from meetings of the Internet Engineering Task Force (IETF) to a remote audience that has expanded from 20 to over 600, and led to the establishment of the Multicast Backbone (MBONE). The MBONE is a virtual network overlaid on portions of the physical Internet to provide IP multicast connectivity among the participating sites.

For this demonstration, workstations at MM'94 will be connected to the MBONE and will be able to teleconference with locations around the world using a collection of free software tools for session control and for audio, video, image and shared whiteboard media. These tools, which have become known as "the MBONE tools," were developed by several organizations and have been ported to DEC, HP, SGI and Sun platforms. MM'94 attendees are encouraged to learn how their home institutions can connect to the MBONE and join in this experiment.

For now, audio and video on the MBONE get the same best-effort datagram service as other kinds of traffic on the Internet. This is only satisfactory when the load is light. The second half of this demonstration will show how the low delay required by real-time media can be assured through resource reservations. Researchers on DARTnet, an ARPA-sponsored testbed, and in IETF working groups have developed traffic control mechanisms for IP routers to provide privileged service to real-time traffic, plus a resource reservation protocol (RSVP) to allow applications to specify their quality of service requirements. In this part of the demo, real-time audio and video will be sent while competing test traffic is generated to overload the network links. With reserved resources for the audio and video, observers will see the teleconference proceed without interruption, while without a reservation, the audio and video will break up badly.

11 Distributed Music Trio – Media Synchronization Over Wide Area Networks

Julio Escobar
Systems and Technologies Division, Bolt Beranek and Newman, Inc.

Contact: *jescobar@bbn.com*

Synopsis: Multimedia synchronization is a subject of major interest within multimedia communications. This demonstration will show the ability to display (play out) a music orchestra at the conference site by remotely synchronizing the audio and video from musicians across the country. It will serve to demonstrate the current capabilities of multimedia synchronization.

The demonstration will synchronize across the continental USA two live players and one repeater site, thus delivering a synchronized music trio performance to the demo site in spite of delay variations through the internetworks that carry the audio and video flows.

The demonstration will also include Graphic User Interface displays of the measured and managed delays in the system to show in real-time the way in which the synchronization protocol operates. The audience will be able to experience the difference between synchronized and unsynchronized music performances.

12 CMIFed: A Transportable Hypermedia Authoring System

Guido van Rossum, Lynda Hardman, Jack Jansen, and Sjoerd Mullender
Centrum voor Wiskunde en Informatica (CWI)

Contact: *Guido.van.Rossum@cwi.nl*

Synopsis: CMIFed is a prototype of a novel authoring environment for transportable hypermedia presentations (interactive presentations including text, graphics, video and sound, as well as hyperlinks).

The author of a presentation mainly uses two "views" on the presentation: a hierarchical view showing the presentation's logical structure, and a time-based view showing the timing structure as well as the assignments of media data objects to logical "channels". The latter combine aspects of resources, such as screen real estate, with style definitions, specifying attributes like font or color for media data objects assigned to them.

The demonstration shows how a hypermedia document can be created, modified and previewed using CMIFed.