Integrated Media Systems:
Toward the Media Immersion Environment

By
Dennis McLeod, Ulrich Neumann, Chrysostomos L. Nikias and Alexander A. Sawchuk

Integrated Media Systems Center
School of Engineering
University of Southern California
3740 McClintock Ave.
Los Angeles, CA 90089-2564
Telephone: (213) 740-0877/Fax: (213) 740-8931
nikias@imsc.usc.edu

1. Introduction

The Media Immersion Environment (MIE), a vehicle for the system integration of multimedia engineering and other scientific developments, is an evolutionary platform that has a robust system infrastructure for the implementation and demonstration of multiple, diverse system applications in a simulated operational environment. An evolutionary platform that will change dynamically over time, the MIE serves as a national testbed for the integration, evaluation and demonstration of advanced multimedia technologies. Inherently multidisciplinary, the MIE encompasses all media systems integration efforts to create powerful, advanced media systems.

Today, the integrated media system, a computer-based facility that supports the creation, usage, sharing, distribution and effective communication of multi-modal information across the boundaries of time and space, is rising to the forefront as the information technology centering point for the next decade, expanding the way we communicate, work and conduct business.

The Media Immersion Environment provides a framework for the creation of various integrated media systems, a framework that offers an overarching, unifying theme for "fitting together" the different information technologies. In an integrated media system, advanced media technologies are used to combine, deliver and transform information by way of images, video, audio, animation, graphics and text in real time.

In the near-term, it is expected that the following functionality will be supported with the MIE:

- real-time video and spatial audio with head tracking;
- storage and access to immersipresence sessions;
- connectivity to Internet database resources; and
- high speed networks and interfaces with customized protocols.

The mid-term MIE will go several steps further, providing:

- 3D sensing and representation of people, location and activity;
- user customized information selection and presentation; and
- sustained high bandwidth Internet-II connectivity.

Finally, in the far-term, it is anticipated that the MIE will provide:

- portable media connectivity;
- information customization to person, environment and context; and

- multi-modal visual, aural and haptic interfaces.

The Media Immersion Environment can provide a number of key unique features, including:

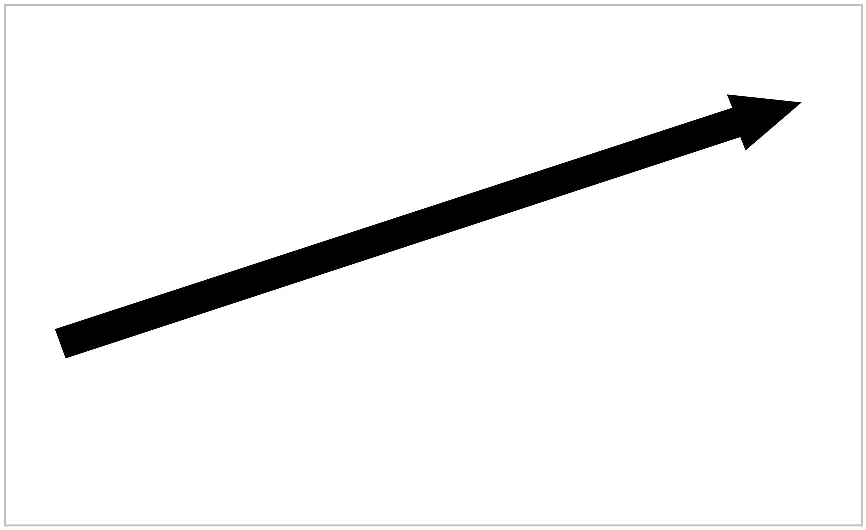
- explore high levels of integration and functionality;
- target low-cost, scalable, very high-performance solutions across boundaries of time and space;
- leverage multiple research efforts in multimedia with different time scales for deliverables;
- investigate the integration of a broad range of technologies with interoperability between major system components;
- be a national test environment with long-term continuing value;
- target a wider range of technology integration with deliverables phased over multiple years; and
- encourage non-proprietary, vendor independent solutions.

The concept of an immersive telepresence system—a new kind of teleconferencing—presents an example of an application being developed within the MIE framework. In the concept, meeting attendees in different buildings will appear and sound as if they are in the same room. (Figure 1) This immersive telepresence system will support the real-time delivery of uncompressed video and spatially-rendered audio, and it will be capable of both real-time access to the content of these sessions and on-line storage of the sessions. Such a teleconferencing system would represent a major step in the evolution of both desktop video conferencing and room video conferencing. Desktop video conferencing moved from the use of ISDN and LAN connections in 1991 to the use of standard telephone lines, ISDN and the Internet in 1997. Room video conferencing evolved from proprietary systems in 1991 to PC-based systems in 1997. As part of the Media Immersion Environment, the immersive telepresence system will be richly interactive, scalable, fully immersive and connective. (Figure 2)

Various integrated media systems will be developed within the MIE framework to give the individual the same kind of multimedia immersive experience as the immersive telepresence system. This experience can be described as cooperative immersipresence in which an individual is immersed in a controlled, customized multimedia information universe, a universe that will

Fig. 1 - A concept of collaborative interaction between two groups in different buildings using an immersive telepresence system.

include people, places, information collections, databases, the web and beyond.



In achieving cooperative immersipresence, the integrated media system generates sensations in humans that are comparable to those generated by the real world. Such as system provides the seamless integration of "atoms" with "bits" and augments the real-world environment with the virtual world. The developmental stages of cooperative immersipresence include creation, access, dissemination and interative use. Specifically, cooperative immersipresence is achieved through the use of rich multi-modal (visual, auditory, haptic) interface techniques to support the creation and use of multimedia information (the Computer Interfaces research area). Multimedia database structuring, indexing, and management techniques are employed to support information storage, access, sharing, and customization (the Information Management research area). High-performance wired and wireless communication, transmission, and compression techniques and mechanisms are used in dissemination and to support multimedia information immediacy and portability (the Media Communications research area). Demonstration Technologies showcase the interactive use of multimedia information. (Figure 3)

The advancement of integrated media systems in developing the cooperative immersipresence experience requires innovations and system integration in three key areas. First, the systems must seamlessly combine digital video, audio, text, animation, graphics and knowledge about such information units and their interrelationships. Second, they must integrated with humans, via cooperatively interactive multi-modal interfaces. And third, integrated media systems must communicate with other such systems and content-addressable multimedia databases, both logically (information sharing) and physically (information networking, compression, delivery). There is a potent emerging market for integrated media systems. Use of the Internet and the World Wide Web has exploded in the past several years, increasing from 9 million hosts in 1995 to 81 million in 1998. (Figure 4)

Entertainment, manufacturing industry and education are just some of the sectors that can benefit from the merger of technology and content through use of integrated media systems. For example, visualizations of data and virtual systems in the context of the real world can make information spatially indexed and more understandable. Figure 5 illustrates a movie director viewing a sound stage that contains a virtual model of a stage-prop under construction. From this augmented reality visualization, the director can make better decisions about the details of the prop, the staging of scenes around the prop, and the financial impacts of the decisions. This example is typical of many similar future applications that require advances in displays, wide are tracking, wireless networks, distributed information sharing, content-based addressing, and continuous media delivery.

Industrial workers often require instructions and guidance in performing maintenance and manufacturing tasks. The time spent searching for information and clarifying conflicting data

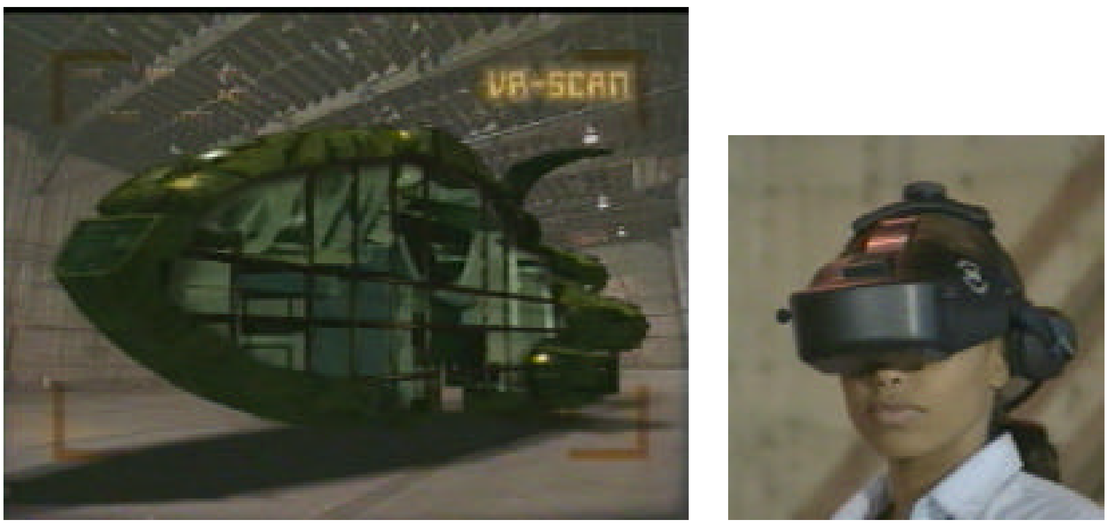


Fig. 5 – A movie director interactively visualizes and modifies a virtual model in the context of the real sound stage before the prop is fabricated and scenes are shot.

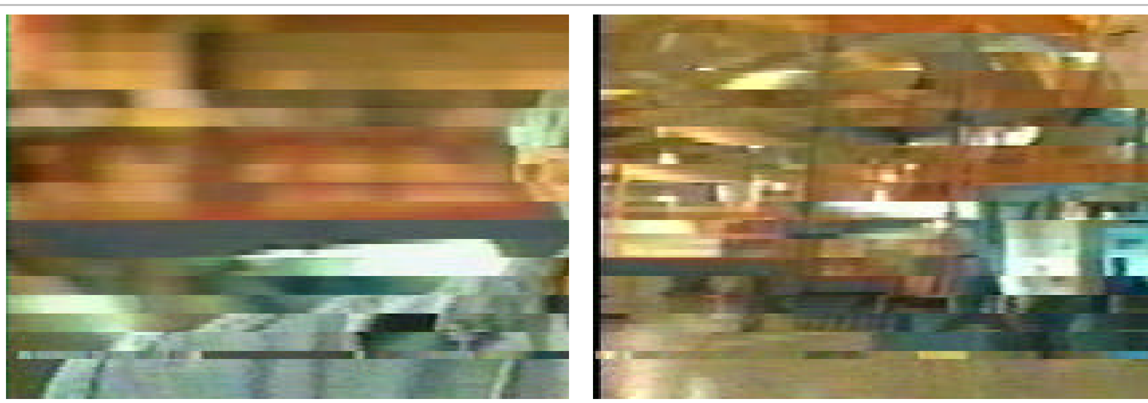


Fig. 6 – Confusion about a manufacturing task is resolved with a remote expert by communication of speech, gestures, 3D scene representations, and database consultation. A representation of the remote expert appears to the worker when he uses the glasses.

reduce productivity and increase the potential for errors. Figure 6 illustrates a worker obtaining clarification of a task detail from a colleague or expert at a remote location. Rather than engaging solely on a verbal description of the scene and problem, advanced technology produces a 3D image of the people and their surroundings. With this capability, the problems are more naturally and clearly communicated and resolved as if the remote expert was physically there. Furthermore, automatic logging and analysis of the problem-solving session updates the process and instruction database. Relevant technologies for this scenario are low latency wired and wireless networks; automatic 3D scene synthesis and motion tracking; animation from sparse motion data, audio analysis and speech recognition; and content-based information addressing.

The rapid pace of scientific discovery and development provides new challenges for education. Current and future scientists and engineers need the ability to understand and synthesize solutions to problems that they were never trained to grapple with. Interactive information systems can facilitate a continuous career-long learning process through information access and learning assessment coupled with experiential discovery facilitated by experimentation and simulation. Figure 7 illustrates a future classroom where groups of students interact with simulations and visualization of biological processes. Distributed systems and instructors interact with the simulation to present new challenges and reinforce difficult concepts while monitoring the class progress. New tools and approaches to content authoring and modeling for simulation are required to realize these systems. New models of learning and assessment are also required to measure student progress and guide them through a learning and problem solving process.

These examples of cooperative immersipresence are portals for individualized access to a universe of heterogeneous information sources. An individual is immersed in a controlled, customized multimedia information universe, which includes people, places, information collections, databases, the web and beyond. The cooperative immersipresence portal serves as an extension of an individual's memory, supports information discovery and understanding and enables interactive guidance, learning, exploration, entertainment and interpersonal collaboration through natural and familiar human metaphors.

2. The Integrated Media Systems Field

Because of its deeply cross-disciplinary nature, multimedia's scientific history can be traced back to a number of distinct yet interrelated fields in engineering as well as related disciplines. Clearly,

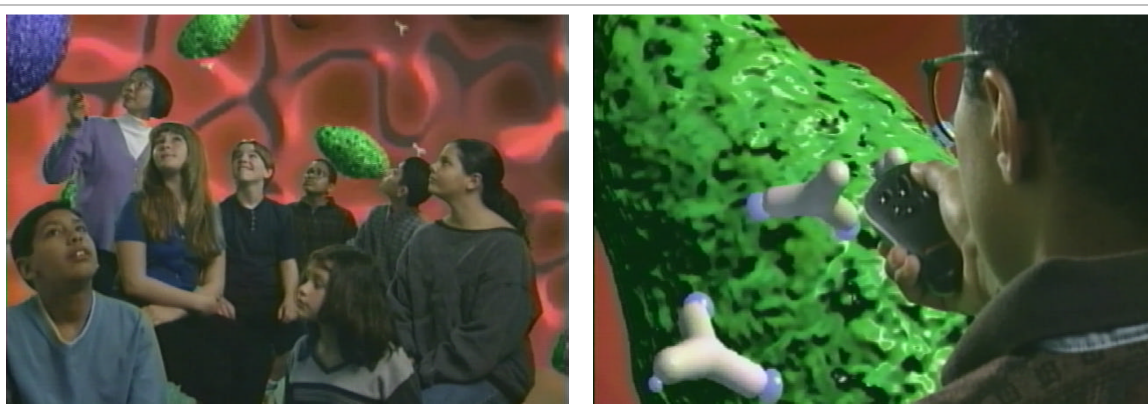


Fig. 7 – Students interact with a biological cell simulation to explore and understand the behavior of the immune system.

Turing's and von Neumann's research conducted in 1930s and 1940s on the theory of computation laid out the groundwork for computer systems used today. Also, Claude Shannon's work in information and communications theory and Wiener's work in filtering and prediction theory provided the bits and bytes building blocks of today's computing and communication systems.

In addition, today's burgeoning multimedia research efforts would not have been feasible without research that led to hardware and software tools such as the computing power of VLSI microprocessors and multiple chip modules which enable device manufacturers to package an ever increasing number of transistors into single miniature packages complemented with adaptive operating systems capable of incorporating a diverse set of specialized system software tools. The complex processing done today on personal computers could only have been done on supercomputers just a few years ago.

An illustration of how far the technology has moved in the last decades can be found in interactive media. The demand for interaction began long before technology could support meaningful two-way communication via a mass medium. Witness, for example, the popularity of network television's first "interactive" show, a CBS children's program called *Winky Dink and You* in 1953. Viewers were able to "interact" in creating their own medium by covering the television screen with a clear plastic sheet and then connecting dots. (The show was canceled after children began drawing directly on the screen instead of buying the plastic sheet.) More recently, mass media companies pursued new modes of two-way communication, such as the Warner Amex Qube experiments on cable (1970s), the 1979 introduction of the first videotext experiment in England, and the commercial failures of interactive video-based news and commercial services in Los Angeles and Miami in the 1980s. Although many factors, including high cost, limited content and marketing problems, contributed to the demise of these services, it is also clear that the technology of the times did not support the demands of the user for the access, presentation and interactivity promised in these trials and did not address the human/computer interface issues.

Even the 1969 launch of ARPAnet and the spread of the Internet and Bitnet had limited impact before the introduction in 1989 of the World Wide Web and user interfaces to it. Since then, the technological developments in video and audio compression, indexing, retrieval of information, intuitive user interfaces, and a long list of other breakthroughs have resulted in the interactive multimedia method of communication. It has also brought together researchers and creators from a spectrum of disciplines - science and engineering, communication, music, cinema, television, education, library science and the arts.

In the computer interfaces area, important research includes Ivan Sutherland's work in the 1960s in inventing graphics-based interactive computer systems. Related to Sutherland's research is Tufte's work in the 1980s in information visualization, which has helped investigators understand how humans look at a complex object (e.g., a weather map) and, at a glance, are able to understand a number of complicated concepts about it. The application of semiotic theories to user interface and icon design (such as the work done by Xerox, which led to Apple's introduction of the Lisa computer) resulted in today's "user friendly" computers. Also significant is the research done regarding vision and recognition in which computers are tasked with interpreting images. The image understanding and, more specifically, the object recognition problem are long-standing and difficult. Success here has largely been achieved by either focusing on specific kinds of images or application domains. The more general object identification and understanding problems are still open.

Engineering and technology developments resulting from this graphics-related research include off-the-shelf graphics hardware that allows users to easily model and render images using personal computers. Another significant development still being honed is real-time image processing. This technology has moved into the marketplace in a number of areas, including 3-D computer-aided design, digital special effects and digital video and audio.

The state-of-the-art in multimedia information/data management can be traced back to foundational research in the 1970s in which the relational database model was introduced by E. F. Codd and associates. This was a major breakthrough in focusing on organizing information in logical ways appropriate for users, and separating this from internal organization and accessing details. In the 1980s, research focused on developing semantic data models, which concentrated on determining new ways of describing information with a higher level of abstraction and emphasis on inter-relationships. In the 1990s, investigators have pushed database management research even further as they have pursued object-based databases and extensible relational systems, which will allow users a more concept-based method of archiving, sharing and searching for information.

Also key in the information management area are technological developments in the use of hierarchical storage systems and parallelism for scaleable, real-time data access to large databases. The development of "federated" database systems, which balance centralization and autonomy/ distribution, has also been central.

In the media communications area, perhaps one of the most significant moments of scientific understanding arrived in the 1960s with the development of the laser by the Los Angeles-based Hughes Research Labs. Harnessing light into a high-density source coherent over a wide range of frequencies has revolutionized the scientific community, including work in astronomy, life sciences, communications and medicine. Research in the physics of optical propagation provided ways to work with lightwave technology in different media, including fiber optic technology, which is particularly relevant to multimedia research and applications.

Also key has been the pioneering research conducted in coding and spread-spectrum which has advanced efficient signal design, coding transmission and manipulation. During the 1960s, USC investigators dramatically advanced digital communications science through their research on pseudonoise generation and synchronization systems. This research, which continues through USC's communication group, has successfully pioneered the research and provided image processing services over the application of CDMA spread-spectrum communications to cellular services--a vital component to multimedia delivery. Also beginning in the 1960s, data compression technology research has led to such developments as high definition TV and digital video over the Internet These developments have made, and are continuing to make, an enormous impact on the consumer market, including CD-ROMs, mobile telephony, personal computers, DSS, Web TV and the Internet.

The integrated media systems revolution has been launched due in large measure to our ability to transform and integrate signals of all kinds into digital representations which, in turn, allows easier storage and manipulation. The impact of digital signal processing research in filter design and implementation as well as advances in spectral analysis over the past 30 years have guided the way for today's robust interference mitigation and classification methods which have cut across the boundaries of time and space.

3. Research and Development in Integrated Media Systems

In defining and structuring a research and development program, one needs to analyze the current state of science and technology for integrated media systems and then define a research program that will dramatically advance integrated media system performance, features and functionality. In sum, there are four key top-level technical barriers: (1) computer interfaces that are predominately unidirectional, inefficient and of limited effectiveness in user communication; (2) real-time distribution and storage of multimedia information that is expensive and slow; (3) an effective methodology for managing, accessing, understanding and sharing integrated or distributed multimedia databases does not exist; and (4) the production and testing of integrated media products for education, training, manufacturing, the arts, mass communication and entertainment are both costly and time consuming.

Accordingly, the research goals should focus on surmounting these key technical barriers. One should seek to: (1) develop perceptive interfaces for truly interactive media systems (for creator-computer and computer-consumer interfaces); (2) provide high-bandwidth information delivery with real-time artifact-free (relative to human perception) compression and decompression; (3) provide effective content indexing and access to distributed, heterogeneous multimedia collections/databases, as well as the ability to fuse, understand and integrate information from multiple sources, and (4) devise low-cost end-user integrated media systems to demonstrate integrated media systems technology in the areas of education/training, mass communication and entertainment, virtual (and "tele") presence and scientific collaboration.

The Research Areas and their component projects--Computer Interfaces (CI), Information Management (IM), and Media Communications (MC)--should be at the enabling technologies level of the research program. The underlying basic science should be below this level. The Media Immersion Environment (MIE) should be atop the enabling technologies. It is through the MIE that substantive interdisciplinary and multidisciplinary solutions are integrated into a unified framework that is domain-independent and coordinates with industry and open architecture standards. Also atop the enabling technologies should be Demonstration Technologies (DTs). Examples of DTs include BioSIGHT, Personal AudioCast, the Interactive Art Museum, and the National Tele-Immersion Initiative (NTI). These Demonstration Technologies (described further below) serve to support specific experiments and demonstrable prototype developments in key areas of application.

3.1 Computer Interfaces

New high-bandwidth visual and aural interfaces and metaphors are needed to replace (or enhance) today's low-bandwidth keyboards and mice. Audio, speech, haptics, video, and 3D graphics will drive, and be driven by, new applications in which elements will be merged and manipulated interactively. The vision of the Computer Interfaces (CI) research area is to expand the diversity and effectiveness of the communication that transpires between people and computing systems. We consider three directions critical for impacting future computer interfaces:

- expanding the range of communication modalities and capabilities;
- migrating interaction from the desktop to more human (natural) terms; and
- seamlessly manipulating , fusing, and understanding real and virtual sensory information.

Our vision of the "ultimate" interface is not any single widget or device; rather, it is a collection of many engaging, subtle and communicative sensory interactions. Visual, auditory and haptic interactions and sensations are dominant communication modalities due to their familiarity and rich expressiveness. An interface based on these modalities should be transparent to the user, requiring no special physical effort or mental translation to facilitate communication. Abstractly, an appropriate analogy is that of a highly trained and attendant human assistant responding to a user's gestures, pose, voice, gaze, expression, and keystrokes - perhaps even anticipating them. At a semantic level, the interface is capable of sensing, synthesizing, manipulating and presenting visual, aural and haptic sensations, subject to the user's preferences and complete control.

To progress toward the vision of the ultimate interface, substantial breakthroughs are needed in basic science and engineering. Sensors must be improved to capture the fidelity of real-world stimuli, and transducers (displays) are needed to render both sampled and synthesized sensations. Models and representations of sources, environmental effects, sensors, displays and human perception are basic scientific issues to explore and develop for aural, visual and haptic sensations. Intelligent interface behavior requires recognition and interpretation of representation streams.

Technology must be conceived and efficiently engineered to provide the immediacy needed in interactive endeavors. Metaphors must be developed and applied within applications to create symbiosis in the complex system of heterogeneous interface elements and users working to achieve their purposes.

The above barriers to this vision can be distilled into three terse fundamental knowledge themes (subareas). Visual sensing, display and tracking deal with basic syntactic input and output capabilities such as image acquisition, representation for display and tracking of pose and object motions. The theme of scene analysis, modeling and representation deals with the semantic issues of identifying the objects in the scene, extracting models of objects and representing the scene as geometry, textures and procedures for rendering. Aural rendering and speech processing comprise the third knowledge theme. Both spatial audio and speech recognition methods build on perception models, so they are integral to the theme. A fourth knowledge theme of human factors, perception and evaluation is implicit.

Visual Sensing, Display, and Tracking

An augmented reality (AR) creates the perception of computer-generated information fused with objects and spaces in the real world. This novel display and interaction metaphor breaks the barrier between the desktop monitor and real world. Tracking has been at the heart of research and development in augmented reality since its inception in the 1960s. Tracking systems employ a variety of sensing technologies, each with unique strengths and weaknesses, to measure the six-degree of freedom pose of an object or person. An important research goal is the development of robust methods for (object-centric) tracking, suitable for AR applications in indoor industrial environments. It is necessary to focus on overcoming several historical weaknesses of vision-only tracking systems. A Kalman filter approach has been developed to dynamically add features to the world model known by the system, thereby expanding the range of tracked viewpoints. General recognition of objects and features is unreliable, so fiducials have been designed that are inexpensive, passive, non-invasive and easily applied. Methods for recognizing the subset of visible fiducials are being developed. These cluster recognition methods must be robust, fast and extendible at run time.

Scene Analysis, Modeling and Representation

Interaction with 3-D graphics models is difficult to extend to real world scenes. Modeling is tedious and limited. Even when models are available, the complexity of the real world is beyond the capabilities of current representations. It is necessary, therefore, to turn to images of a real scene as the representation for interactions. A research goal is to develop novel algorithms for synthesizing new views of a static scene from multiple pictures of the same scene. Both pure image-based and reconstruction based approaches have been explored. The pure image-based approach currently only solves the problem in special cases such as parallel images or multiple views. The pure reconstruction-based methods require fully calibrated cameras. Both shortcomings are overcome in this approach. In some cases, it is necessary to create 3-D models of real objects. The goal is to produce a 3-D model of an object from a single image, with a minimum of interaction. A user provides limited but pertinent, high-level information to a robust algorithm that extracts geometric and texture representations automatically. The computed models should be of computer graphics quality, i.e. suitable for photorealistic rendering. The user interaction should be intuitive and simple, while most of the modeling task is performed automatically.

Aural Rendering and Speech Processing

Numerous applications are envisioned for integrated media workstations that support functions to manipulate, edit and display still images, video and computer animation and graphics. The necessity, however, to accurately monitor the sound associated with visual images created and edited in the desktop environment has only recently been recognized. This is largely due to the increased use of digital audio workstations that have benefited from rapid advances both in main

CPU computational power, as well as in special-purpose digital signal processors. In addition to accurate reproduction of the measurable characteristics of sound (e.g., frequency response and dynamic range), multichannel and emerging 3-D audio program material requires accurate spatial perception of sound as well in order to create a seamless aural environment and achieve sound localization relative to visual images. A mismatch between the aurally-perceived and visually-observed positions of a particular sound causes a cognitive dissonance that can seriously limit the desired suspension of disbelief. Research is being focused on system implementations that address these needs, initially for a single listener, and in the long-term for multiple participants.

Robust speech recognition offers a natural means of human-machine communication. A key challenge lies in the proper handling of complex background noise. Novel approaches to major modules of the recognition system are being explored. In the dynamic synapse approach, temporal pattern matching requires adaptive learning and massively parallel computation. The parallel processing nature of the neural network can achieve high-speed operation and good accuracy if properly implemented in the silicon hardware. Detailed analysis and system-level simulation are performed to identify parameter values for silicon microchip implementations. A four-channel continuous interleaved sampling speech pre-processor for cochlear implant devices has been developed that will help people who are profoundly hearing impaired. Low power design methods are used to reduce the power consumption for the portable electronics that have become very important in multimedia systems.

Each of the three sub-areas of CI research is critical to the Media Immersion Environment (MIE), with the MIE providing a platform for the demonstration of CI developments. At the same time, requirements generated by the Technology Demonstration and the MIE integration process feed back to influence the research program and confirm or dispute current scientific theory. The MIE computer interfaces should evolve through approximately three distinct stages. The initial MIE will feature high quality video and head-tracked immersive audio. In mid-term, real-time scanned and rendered 3-D scene representations supplanted video images as the dominant visual media. The long-term MIE goal is to provide a complete multi-modal suite of interfaces that approach the capabilities needed in the vision of the ultimate interface.

3.2 Information Management

The Information Management (IM) research vision is an environment in which distributed/federated, heterogeneous, multimedia information can be accessed, retrieved and presented, content-indexed for a diverse spectrum of users in various applications domains. In this environment, a multi-level sharing network (federation) supports the discovery of remote information and the browsing of the distributed heterogeneous information space, and allows users to integrate and fuse information from multiple sources into a local “information space” and share information selectively with others. People, places and things are virtually incorporated into the information space(s), as a foundation for cooperative immersipresence.

In the Information Management research area, emphasis is placed on addressing three key technical barriers to providing effective multimedia information access and retrieval, presentation, content indexing, sharing and exchange and robust, seamless delivery. First, the heterogeneity of multimedia information at various levels (content description, format, storage, modality, transmission) must be accommodated, supporting translation, and integration/fusion of data from multiple sources. Second, more effective information communication between user and information collections is required, to support customized selection, presentation, and sharing capabilities. And third, scalable techniques and mechanisms are required for seamless, disturbance-free storage, compression and access to distributed multimedia information repositories/libraries. In particular, effective techniques for disturbance-free delivery of video and audio are lacking.

Based upon these key technical barriers, the research program in the IM area encompasses four cornerstones of research technology to support distributed multimedia information management:

- content representation and extraction (content indexing),
- integration and fusion (of distributed, heterogeneous data),
- user-customized selection and presentation, and
- storage and access for continuous media (video, audio).

These themes correspond to the sub-areas of enabling technology research within the IM research area. Summarized immediately below are the key aspects of work in each of these sub-areas, with emphasis on the integration of specific work with the MIE and with key efforts in the CI and MC areas.

Content Representation and Extraction
The main objective here is to develop techniques and paradigms to both extract information from multimedia data types (e.g., audio, image) as well as represent the extracted information in a way that the corresponding data can be queried later. The content-based indexing and retrieval of multimedia information can be supported by utilizing these techniques within a common data model and representation language. Each multimedia data type has its own specific characteristics and hence techniques developed to extract information from one media type cannot be directly applied for content extraction from another media type. For example, two different sets of expertise are required to develop tools for content extraction from image data as opposed to semi-structured text data. While the first one relies heavily on signal and image processing expertise, the second one utilizes learning and artificial intelligence paradigms. Even for a single media type such as image data, extracting (say) color features requires different techniques as compared to those required for (say) extracting texture features. Finally, many real-world applications query multimedia data based on neither a single feature (e.g., color of an image) nor a single media (e.g., only audio part of a movie). Therefore, an integrated system is required with all the content-extraction algorithms at its disposal. In addition, the system should be able to represent the content of multimedia information in a unified manner, independent of the content extraction algorithms, so that a human user can reason about this content. Such a system is being developed within the context of the Media Immersion Environment.

Integration and Fusion
One goal is to develop a "multimedia database network" that supports navigation, sharing, and integration of structured multimedia information. Using a common interchange language, this facility will support the sharing and integration of information. Three general problems are being addressed to achieve these objectives. First, the languages and ontologies are being developed for representing the various multimedia information sources. This includes the techniques and tools for reconciling representational difference between different sources. Second, the query processing capabilities are being constructed to efficiently support access to a wide range of multimedia information sources. Unlike traditional relational databases, access to multimedia data can be very costly. Third, in order to integrate multimedia data, techniques are being devised for combining different types of multimedia information, e.g., combining semi-structured text documents with images containing maps or photographs.

User-Customized Selection and Presentation
The overall objective here is to devise and demonstrate general-purpose, scaleable techniques and mechanisms for user-customized selection and presentation of multimedia information. A key goal is to focus on techniques essential to the MIE, and in particular the information space access portion of the MIE. To this end, present research stresses the analysis and modeling of the

communication effectiveness of alternative presentations. The information space subcomponent of the MIE will present multi-source, multi-type and multi-format information ("multimedia") such that the user will receive it in a manner unhindered by that user's pre-existing expectations. It is necessary to focus on a delivery system in which the range, format and style of presentation is customized based on a number of variables (e.g., age, gender, cultural background, etc.) in addition to the user's preferred order of presentation of written material, and potentially of the writing style itself.

Storage and Access for Continuous Media
In addition to traditional data types (e.g., text, images, etc.), a multimedia database management system must support continuous media, audio and video clips. The principal characteristic of continuous media is their sustained bit rate requirement. If a system delivers a video or audio clip at a rate lower than its pre-specified rate without special precautions (e.g., pre-fetching), the user might observe frequent disruptions and delays with video and random noises with audio. These artifacts are collectively termed hiccups. The focus of this effort is to develop hiccup-free display techniques in support of continuous media. Centralized servers that can scale to support thousands of clients and geographically distributed systems that consist of many such servers have been investigated. Techniques should be scalable, and the system should be designed to grow incrementally as the requirements of an application increase in order to avoid a degradation in performance. For distributed servers, the focus is being placed on techniques that allow multiple servers to communicate and collaborate in order to stream many simultaneous displays to geographically distributed users. The main focus is to avoid any of the system resources (e.g., servers, links) to become either a bottleneck or a single point of failure. A number of these algorithms have been realized in the prototype system called "Mitra."

From a scientific point of view, research in the IM area focuses on four key aspects of fundamental multimedia information management knowledge. First, a theory should be developed of multimedia information unit content representation, and multi-dimensional content extraction from imaging, audio, video, and text. Second, a model should be devised of multimedia, multi-source information integration and fusion, with a knowledge-based aspect for semantic and ontological heterogeneity resolution. Third, a model and techniques for multi-perspective, multi-cultural information structuring and design should be developed, including an experimentally analytic aspect (human factors perspective). Finally, architectural principles, benchmarking, and evaluative techniques should be devised for scalable storage and delivery of continuous media information (particularly video and audio).

The Media Immersion Environment (MIE) provides a platform for the demonstration of IM developments in all four of the subareas above. At the same time, requirements generated by the Technology Demonstration and the MIE integration process feed back to influence the research program and confirm or dispute basic scientific theory. The MIE multimedia information management capabilities should evolve in three stages. The initial MIE will feature storage of, access to, and editing of immersipresence sessions; connectivity to Internet-based information sources will also be supported, including the fusion of such information into the immersipresence environment. In mid-term, robust user-customized information selection and presentation facilities will be provided, along with more functional information sources connectivity. In the long-term MIE, information customization for person, environment, and context will be supported

3.3 Media Communications

In the Media Communications (MC) research area, the long-range strategy should be to provide the innovative, forward-looking wired and wireless communications that serve as the backbone for advances in media interfaces, information storage and access and user applications.

The prime problem for today's technology is the low data bandwidth available to individual multimedia users. Data rates of several hundreds of megabits per second are needed for real-time archival quality digital video at current resolution. New multimedia environments such as collaborative telepresence, personal augmented reality and high-definition digital TV may require even higher data rates. Even though auxiliary coding and decoding technologies such as video compression may reduce these rates by one or two orders of magnitude for some applications, the resulting data rate of a few megabits per second is many times faster than services available to most businesses and individual consumers. Intimately related to these issues are problems in the storing of thousands of terabits of information, the management and low-delay transmission of this data over distributed shared networks, the signal processing and communication technology required to implement high bandwidth wireless transmitters and receivers that are robust in noise and interference environments and the complexity and power consumption of the VLSI hardware needed to implement these functions.

These issues are summarized by the four main areas of enabling technology that are central to our media communications vision:

- video and audio compression and networks,
- high bandwidth multiplexed wireless,
- photonic networks, systems and devices and
- digital video networks and interfaces.

The following sections summarize the key aspects of work required in each of these sub-areas, with emphasis on the integration of specific work with the MIE and with key efforts in the CI and MC research areas.

Video and audio compression and networks

Achieving the long-range goals of integrated media systems as outlined in this vision will require high quality digital video to be transmitted over limited bandwidth communication links. This demands advanced compression using object features and new mathematical signal representations. Advanced sensory experiences such as 3-D television will require even greater bandwidth, more complex compression, and completely new procedures for distributed data retrieval and caching over networks. Thus, this area of research has close interactions with the Information Management (IM) research area and the high-speed network projects within the Media Communications (MC) area. Some specific projects within this sub-area include: algorithms for structuring compressed bit streams for ease of data manipulation, including scalability (transmission over heterogeneous networks, retrieval from a database), editability, and error resiliency; techniques for post-decompression enhancement having a resultant image quality that is consistent with human-observer quality perception; and using embedded wavelet transforms and multi-resolution motion representations to improve the temporal, spatial and quality scalability of compressed video.

High bandwidth multiplexed wireless

Many aspects of developing multimedia technologies need wireless communication technology with much greater bandwidth, lower power operation, and robust connections that are free from multipath fading and shadowing. Today's wireless systems have low data rates (< 200 Kb/s) and require high radiated power (100 mW). With non-robust, lossy channel conditions, video transmission is poor or impossible. Improvements in bandwidth (to > 1.5 Mb/s) with lower power (< 1 microwatt) will be achievable only with completely new approaches such as high bandwidth impulse radio. These techniques increase transmission bandwidth using spread-spectrum pulse-hopping or impulse radio techniques to resolve multipath and reduce the power requirements. Thus, this work involves low power VLSI design, multipath radio-frequency signal modeling and measurement, signal design, signal processing, high speed synchronization and interference mitigation technology. With the results of this work and the use of adaptive source rate encoding and error correction, full-motion robust wireless video should be possible in the future.

Photonic networks, systems and devices
In the future, both photonic and electronic technologies will be utilized for high bandwidth communication and computationally intensive signal processing of multimedia information. Optics is advantageous for long distance fiber transmission, for performing the high speed switching functions in networks and parallel processors, or transferring many signals in parallel on or off a chip. Electronics excels at signal processing, switching and logic operations. Communication over fiber optic channels will require efficient interface and conversion between signals in the time, space and wavelength domains, and high density microprocessors and other VLSI chips will require optoelectronics to provide power-efficient input/output at high bandwidth in a limited area. To address these problems, long-range plans call for research on digital optoelectronic systems that combine optics and electronics to perform high information-capacity parallel interconnections. The smart pixel devices at the heart of these systems combine electronic processing with arrays of light detectors, emitters or modulators on a single substrate. In addition, areas for investigation are optoelectronic devices and architectures capable of efficient, high-bandwidth communications interfaces among time, space, and wavelength domains, and efficient techniques for the use of fiber bandwidth for high performance communication of multimedia information. All these efforts require technologies for optoelectronic integration and packaging that are in their infancy.

Digital video networks and interfaces
In many multimedia applications, very high bandwidth (>1 Gb/s) to each user with very low latency (delay) is essential. Some applications, such as the archiving of motion pictures or medical image processing, must transmit digital video without compression. The long-range research in this area aims at the problems of interconnecting heterogeneous hardware and software to multi-gigabit local-area and wide-area networks; optoelectronic workstation-to-network interfaces that operate at > 1 Gb/s; and high-bandwidth software, hardware and protocol issues of networks used for interactive remote-visualization and video applications.

3.4 Demonstration Technologies

Demonstration Technology (DT) projects can apply enabling technology and MIE results to prototype advanced systems in such domains as education, news and information delivery and telecommunication. Since these domain-specific prototype applications address focused problems, they can be used for evaluating the utility of research efforts and the MIE. All DT projects should drive MIE requirements and provide analysis and insight about the utility, robustness, and effectiveness of the enabling technologies developed. This feedback should arise from task-performance evaluations, human factors studies and perception measurements performed within the DT projects. Examples of DT projects follow.

BioSIGHT
The IMSC BioSIGHT project was established in October 1997 with the purpose of developing a novel methodology that will map a high school biology curriculum into a series of interactive visualization modules. These modules will augment the role of the teacher and will establish the value of student-centered interactivity, both in an individual setting as well as in a collaborative learning environment. Students will be able to interact with the content material, explore new challenges, and perform virtual laboratory simulations. BioSIGHT, in collaboration with TERC of Cambridge, MA, and SRI of Menlo Park, CA, will develop the architecture of the modules that incorporate an innovative pedagogical strategy. The modules include three components: an exploratory challenge that introduces concepts in the form of a puzzle to motivate and engage students; an interactive storyboard that utilizes advanced multimedia tools to interactively convey

content for exploration; and an interactive laboratory in which students can participate in simulated experiments. Learning outcomes will be identified, and the impact on students and teachers will be measured.

Personal AudioCast

The Personal AudioCast (PAC) project will design, implement and test a user-customized audio information-on-demand system. The system will select customized information from a dynamic database and deliver it in three-dimensional audio form. Sports, financial, entertainment and late-breaking news, along with public service information, will be provided. The Personal AudioCast system will introduce a new level of empowerment for news and information consumers, beyond simply allowing them to select stories according to topics or keywords. The PAC system will build custom audiocasts tailored to a user's interests, communications style, past choices and even physical location. To support personalization, a PAC user profile (also dynamic) will specify content communications style and organizational preference characteristics to tailor a periodic or on-demand delivery to the user. Selections from a PAC database will be assembled and packaged for a user as a continuous, three-dimensional audio presentation.

Interactive Art Museum

Central to the mission of USC's Interactive Art Museum project is the development of new technology for enhancing the exhibition of three-dimensional works of art over the Internet. In particular, this project develops and explores technology that allows remote users to interact with the exhibitions. The virtual museum experience will be augmented by the addition of the ability to "touch" priceless museum objects. Haptic acquisition applications will be developed for the inspection and manipulation of three-dimensional art objects. A working demonstration of an application of haptic acquisition and display will be produced through use of Sensable's PHANToM force feedback robotic master and other devices, including the CyberGrasp vibrotactile glove. Usability studies will be conducted to evaluate the effectiveness of the system. Through future research, haptic capability will be enhanced by the removal of various constraints, such as moving away from the table and feeling other properties of objects such as temperature, stickiness and springiness.

National Tele-Immersion Initiative

The National Tele-Immersion Initiative is a project founded to advance the state of the art in tele-collaboration communication technology. A major goal is the creation of communication environment that provides two or more distant users a true sense of being in the same space. Video has inherent limitations as a visual scene representation for person-to-person communication. The lack of eye contact is indicative of the disjointed geometry between the local and remote sites. Many common face-to-face communication cues depend on correct geometric relationships (e.g., head pose, gaze direction and audio direction.) The approach is to abandon video and migrate to three-dimensional representations of a scene and aural field. The 3-D representations of participants will be sensed and displayed within an integrated office system, called the tele-I-desk. The tele-I-desk integrates the functions of a traditional office desk, a computer workstation and a shared-space communication station. Remote sites communicate through high-speed computer networks, such as Internet II. The NTI team is comprised of a number of university and industrial participants. The IMSC members are focused on visual and aural scene modeling and real-time avatar animation and rendering. IMSC industrial sponsors are participating for the visual sensing and the design and fabrication of the tele-I desk. Collaboration with other NTI sites at UNC Chapel Hill and UI Chicago is pursued and coordinated through Advanced Networks & Solutions, Inc. The NTI system is a major driving force behind the MIE, in which major aspects of the network and database research will be integrated.

4. Conclusions

As multimedia technologies have progressed over the past few years, it has become evident that a unifying vehicle would serve to give needed direction to research in the related disciplines. The Media Immersion Environment fills that role as an overarching, unifying framework. The engine within the MIE framework has evolved to become the integrated media system, a computer-based facility powering the convergence of multimedia technologies. And, as the integrated media system develops new, interrelated ways for humans to gather and manipulate information, these new ways fall under the abstract, centering vision of cooperative immersipresence, a controlled, customized multimedia universe. In summary, then, the MIE will serve well as a national testbed for the future work of integrated media systems.