Chapter 18

Future Directions

18.1 Where Are We Today?

Multimedia represents the features and capabilities of a large number of computer components and systems. We have looked at multimedia systems from a scientific point of view focusing on the processing of audio, video and in part, images and graphics. However, since audio, video, graphics, image processing and real-time processing are not new areas, a question needs to be answered:

Is the simple composition of existing concepts, techniques, components and systems a global multimedia solution?

In the following section we provide a few examples of today's state-of-the-art technologies and in so doing, provide an answer to the above question.

18.1.1 User Interface

Until recently, *audio* has been well-known at the user interface level in connection with special speech analysis and synthesis systems. For example, speech analysis applications typically emphasize menu-control, i.e., for hands-free operation.

If audio is going to be part of a multimedia system, audio input and output methods should be dedicated for all applications. Further, the audio I/O (input/output) operations should be connected with known window system paradigms because this connection increases user acceptance. This requirement implies *spatial placement* of the audio output that corresponds to a video window. For example, if a video window is moved to the left, the corresponding sound should also be moved to the left stereo channel. Audio in multimedia systems is as much a part of the user interface as the window systems are of today's workstations.

Hence, current workstations and operating systems need to be modified, and in the next step, new workstation architectures and new operating systems are needed to provide adequate hardware and system support for this simple example of stereo channels and spatial placement of audio.

18.1.2 Operating Systems

Because of their *real-time requirements,* audio and video streams cannot be properly processed with traditional non-real-time operating systems. Moreover, current real-time operating systems satisfy different requirements than multimedia need, i.e., they provide reliability and their end-systems are most often dedicated systems in which the processing paths are known a priori. Multimedia workstations as end-systems do not require such a high degree of reliability, but they need operating systems which should be able to be configured and programmed by the user. Real-time routines shall be plugged in, the way we incorporate device drivers today.

Hence, the current operating systems need to be modified, and in the next step, new operating systems are needed to satisfy multimedia requirements.

18.1.3 Multimedia Documents

Available text processing systems do not embed the attribute *time* as a part of their internal processing.

In multimedia documents, however, the parameter of *time* is an additional component to content and spatial relations. Hence, a multimedia document is more than a mixture of movies and data in a traditional document.

Here, new solutions need to be found which answer questions such as: ***How must multimedia documents be structured to be easily read and how can the authoring process be supported in this way?***

18.1.4 Synchronization

Currently, synchronization among individual data streams is most often implicitly included, for example, in the application or in the compression scheme (e.g., MPEG-2). It is almost always achieved through interleaving or fixed coupling (time or frequency multiplexing) of data units from different media during the transmission or storage process. Television, video recorders and stereo radios exhibit this kind of behavior to achieve synchronization. Additionally, for the most part, only time-dependent data streams are inter-related and synchronized. For example, TV subtitling is an integral part of video (if the subtitle is not displayed through a videotext service).

In a multimedia system, different kinds of streams need to be synchronized because services such as TV, video recorder, stereo radio, etc. will be integrated as the end-devices. Hence, the current mechanisms for synchronization are not sufficient.

New synchronization mechanisms to synchronize all the different media will need to be developed so that, for example, a presentation of a certain image could be coupled with the call of a procedure, implemented by an application developer.

18.1.5 Programming

Further questions need to be answered regarding the *programming* of audio and video processing. Current programming is based on traditional languages such as C, C++, ADA and others. Also, the programming of real-time algorithms for signal processors has been known for a long time.

Programming of multimedia requires structures and control mechanisms so that a programmer can specify continuous types similar to the character type in a text string, and control audio and video flow from the existing environment in a similar way as a character is controlled in a text string.

Hence, new programming languages need to be developed which should include all possibilities known from C, C++ or ADA, but also include new properties such as the handling of continuous media.

18.2 What Are the Next Steps?

All the above examples make clear the necessity of including temporal considerations into existing concepts, techniques, components, and systems. Hence:

The simple composition of existing systems and methods is **not** *a global multimedia solution.*

In the following paragraphs, let us – without certainly discussing all next steps – outline some developments we foresee and believe are crucial:

All currently known multimedia systems (either products or research projects) consider only partial aspects of the time-critical processing of audio and video. Only those critical components that allow for the correct processing of particular multimedia applications and system configurations are implemented. For example, it is always assumed that the system bus capacity is sufficient or that higher priority interrupts will not cause problems during the processing of continuous data. However, interrupts can considerably disrupt the processing of audio and video data because the process management does not include these interrupts in the same way as the other real-time processing. Therefore, resource guarantees are only provided for the most critical resources.

The goal is to integrate all hardware and software components into time-critical processing. This book is intended to be used as a starting point for accomplishing such a challenging goal.

Trends in several areas such as communication systems, operating systems, user interfaces, compression, etc., already indicate the directions, solutions and approaches.

Strong forces, which drive the development of multimedia systems, are multimedia applications. One of the main driving forces is the media entertainment industry. Therefore, based on the example of the Virtual Reality (VR) environment, we show which new challenges are imposed on multimedia systems.

A *Virtual Environment (VE)* can be viewed as a multi-modal, interactive and spatially-oriented user interface, where a *multi-modal user interface* is simply defined as a human/machine interface that actively or purposely fully utilizes interactions and display techniques in multi-sensory modalities (e.g., visual and auditory display, tracking and haptic interface devices). We discuss some of the future directions in the hardware and software components to provide such VEs.

18.2.1 Devices

Major improvements of *hardware* support are expected, because important hardware issues, such as computer performance, mobility and speed, are centers of research and development. This implies that in the future, multimedia devices could be used as mobile end-points. However, the current mobility of multimedia applications is still limited by display, computer architecture, network and tracking technologies. Hence, work needs to be done in these four areas:

- *Displays*

  The trend in displays for VR systems is the HMD (Head-Mounted Display). The most obvious requirement for HMDs is to have a *resolution and field of view* that approaches the fidelity of the human visual system. A display that allows a wide field of view at a low resolution and simultaneous display of a high-resolution inset that follows the user's gaze would be effective. Furthermore, the *depth of field* is also important. Current HMDs all have pixels focused at the same depth. Stereoscopic displays provide depth information using visual cues (stereoscopic images), but present the whole of the images at a single focal length.

  Clearly, HMDs must be made smaller, lighter and non-tethered; and *see-through displays* (see-through HMDs or see-through hand-held displays), used in augmented reality systems, must eventually be constructed so that they allow individual pixels to be transparent or opaque.

- *Processors*

  Workstation processors need to handle real-time manipulation of video streams, real-time creation of textures from raw images, etc. Processing performance is still slow on available multimedia computers. For example, current high-end graphics hardware, such as the Silicon Graphics Reality Engine [Ake93], allow textures to be mapped to 3D surfaces without significant overhead in the graphics pipeline, but creating texture from raw images is slow on these machines.

- *Networks*

  Virtual environments place demands on networks that are similar to those of other multimedia applications. The necessity of low latency will be partially solved by the availability of high-speed networks.

- *Trackers*

  The problems with current tracking systems are ones of *range*, *latency* and *accuracy*. For all systems, low *latency* is important, especially if the user's head is being tracked. The *range* and *accuracy* parameters depend on the particular VR application. For example, desktop VEs do not require a significant range because the user must be able to see the stationary display. On the other hand, *augmented reality systems*, where the user's view of the real world is overlaid with computer-generated graphics, require extremely high accuracy and low latency. These requirements need to be ensured so that the virtual images register properly with the real world.

Further developments can be seen in the provision of devices that *physically interact* with the environment. For example, in the case of a remote manipulation application, if a user grabs an object with a data glove in a simulated environment, a *robot hand* at the remote site may perform the same action on a real object.

Other trends in devices include the implementation of *locomotion interfaces*. Locomotion interfaces can provide the experience of moving around in space while actually being confined in a small space. Examples of locomotion devices are, in addition to *visual displays*, *haptic devices for feet and hands* and *auditory displays.* We briefly describe the haptic and auditory displays. Visual displays were described in Chapter 4.

- *Haptic Displays*

  A ground surface interface (*haptic interface for the feet*) is a device that permits the user to experience the active sensations of walking, running, climbing, etc., in a constrained space. Such a device allows the user to move his or her feet in a natural fashion and provide feedback to the user matching the space-time characteristics of the simulated surface (e.g., an inclined plane). It also senses the behavior (e.g., positions, forces) of the user to control the actions of the device and to provide appropriate information to the other display systems in the VE. Examples of such interfaces are *treadmill displays* and *haptic interfaces for individual feet.*

  A surface pad (*haptic interface for the hand*) is a device that permits the user to read information stored in the computer by experiencing, for example, vibrations from the pad. The computer information is transformed into tactile feedback presented on the output haptic display.

- *Auditory Displays*

  Auditory displays play sound(s) to the user. There are two VE techniques that use this technology. The first technique is where the auditory display presents sounds which exhibit the same spatial and temporal patterns as when listening to sounds in a natural environment. An example is the illusion of self-rotation simulated by a sound field rotating around the head of a stationary subject [Lac77].

  The second technique is cognitive cuing where, for example, visually induced self-motion is augmented by the sound of a virtual wind or engine aircraft [PKG91].

  The trend here is to create a synthesized acoustic field relevant to the illusion of moving through the real acoustic field. Aside from simulating changes in the direction of sound sources (Section 18.1.1), changes in the apparent distance of the sources and changes in the apparent location of the individual within the reflecting environments are important.

18.2.2 Visualization

Real-time 3D visualization is required by certain multimedia applications. Video will play an ever-increasing role in the visualization process at the multimedia interfaces. The integration of video cameras with display devices will allow for tracking people in front of the display.

Using two cameras, stereo vision can be achieved. If the cameras are positioned appropriately for the remote viewer's interocular distance, a reasonable stereo pair of images can be displayed. Alternatively, it may be possible to dynamically construct a 3D model of the visual part of the user's head and transmit this information [OKKT93].

18.2.3 Mobility

Increased multimedia consumption will require the creation of *ubiquitous computing* environments. *Ubiquitous computing* is a term coined by Mark Weiser [Wei91b] to describe a future in which we are surrounded in our everyday life by a multitude of computers so numerous, yet so unremarkable, that we barely know they are there. Hence, a ubiquitous computing environment will consist of a large number of computers integrated into seamless immediate surroundings, connected by high-speed wireless networks.

To provide such an environment, mobile computing and multimedia will have to be integrated. An example of such an application is the playback of stored video on both mobile devices and conventional workstations. A system supporting such an application must provide performance-guaranteed delivery of video data to mobile clients [MSK+93].

The issue which needs to be resolved to create this environment is: how existing connections must be rerouted to maintain current conversations and how various routing databases must be updated to allow new connections to be established when a Mobile Host (MH) moves into the network. There are several algorithms for performing rerouting in connection-oriented networks, for example, *incremental re-establishment* [MSK+93], or *multicast-based re-establishment* [KMS+93], which might be applied to connectionless networks.

18.2.4 Interactivity

VEs require *high frame rates* and *fast response time* because of their inherently interactive nature. The frame rates need to be faster than 10 frames/s and response times must be kept below 0.1 seconds; otherwise, performance severely degrades the illusion of presense.

One major source of delays is the *data access*, which will need to be bounded. For example, in applications which use video servers (Video-On-Demand), the delays in data management come from the seek time to search for data on the disk and from the available bandwidth to move data during the reading process. The trend is to use *Redundant Arrays of Inexpensive Disks (RAIDs)* for efficient storage and retrieval of large amount of data [Lee92]. RAIDs use parallel transfer from an array of drives to support high data transfer rates and high request rates. Their throughput can reach bandwidths of 100 Mbytes/s, although for some VE applications, the necessary data bandwidth is in the range of 300 to 500 Mbytes/s. This implies that, for some applications, the only viable storage medium for data is physical memory, which given the storage requirements, is currently too expensive to be practical (current memories can achieve a capacity of 16 gigabytes).

18.2.5 Operating Systems

Operating Systems (OSs) should support lightweight shared memory processes, thus minimizing the time required for context switching and inter-process communication. The OS should be capable of assuming that high-priority processes can be serviced at very short and regular intervals. Further, developers should be given the capability to determine scheduling through tunable parameters.

18.2.6 Further Issues in Virtual Environments

The main requirements which will need to be satisfied in a virtual environment are *interacting, navigating and communicating* in the visual domain, as well as in the acoustic and haptic domains. Further, *modeling* (geometric, physical and behavioral) and inclusion of hypermedia need to be considered. These requirements set challenges on multimedia interfaces, hypermedia documents and other overlapping areas. Therefore, we briefly discuss the trends in multimedia interfaces and the inclusion of hypermedia into the virtual environment.

18.2.7 Multimedia User Interface

There are several research and development directions which will dominate in multimedia interfaces implementations:

- *Behavioral Modeling*

  Computers will be used more and more to facilitate communication among people. Education and training become more effective and engaging with multimedia interfaces that incorporate video and audio. Hence, behavioral modeling in social settings, which influences the design of multimedia user interfaces, will continue not only in CSCW environments, but also in education and training [Bla94].

- *Intelligent Multimedia Interfaces*

  Multimedia planning based on Artificial Intelligence (AI) techniques holds increasing importance for research. An intelligent interface should provide automation of routine tasks, enhance services, style and clarity. Intelligent interfaces will be essential when we face the prospect of retrieving information from thousands of on-line databases.

- *Using Existing User Interface Paradigms*

  Information we work with is often most effectively displayed using 2D techniques, even in a virtual environment. The most common 2D interfaces use so-called *WIMP (Windows, Icons, Menus, Pointers)* techniques. In the future, ubiquitous computing and virtual environments will provide management of 3D environments; hence, it is important to examine how 2D WIMP techniques can be applied to 3D environments [FDFH92, Lut94, May93a].

- *Personalized Multimedia Services*

  Personalized multimedia services promise a new era in which clients no longer have to search, locate and schedule media presentations. Rather, intelligent *Personal Service Agents (PSA)*, acting on behalf of clients, provide presentations at the client's preferred viewing times [RR94]. Personalized multimedia services are expected to access other services such as storage providers (multimedia servers), network providers and content providers (e.g., publishing houses, news distributors).

18.2.8 Hypermedia

Hypermedia should be embedded in VE because its techniques would allow for linking physical and virtual objects [Mey89]. For example, if a hypermedia node is embedded into the structure of a virtual world, the node can be accessed and audio played or compressed video displayed. This implies that hypermedia capabilities should be available throughout the operating system rather than in separate hypermedia applications.

Another direction in this area is *hyper-navigation*. Hyper-navigation involves the use of nodes as *travel agents* in the virtual terrain. They would allow travel at either accelerated speed in VE or at slow speed using hypermedia links with their node connections.

18.2.9 Multimedia Applications

New VE applications are emerging that present new challenges to multimedia processing.

Tele-robotics and remote manipulation have become more and more important in fields such as hazardous environments, natural disaster management, hospitals, police and fire stations, distributed nursing and education. The basic principles of the tele-robotics and remote manipulation are described in Chapter 17.

Other multimedia applications, where VEs play a part, come from the entertainment area where the development goes toward *location-based entertainment* and *distributed games.*

Location-Based Entertainment

Location-Based Entertainment (LBE) means an intelligent environment for the user's entertainment. LBE's attractions will be the most publicly visible applications of digital graphics, visualization and other disciplines. In the entertainment industry, convincing sensory illusion and natural interactivity can translate into fascination for players and profit for owners. LBEs could be used to help promote cooperative learning by exploiting "the strong web of connections joining commercial play and cultural cohesion" (Nasaw [Har94]).

Generally, LBE will need technology which provides: (1) several degrees of freedom for haptic displays, (2) film frame sizes and HDTV quality for visual displays, (3) VR, (5) cybermotion (computer-driven animated mechanism), (6) mechatronics (electro-mechanical apparatuses), (7) human interface, (8) performance support, (9) automatic translation, (10) CSCW, and (11) groupware.

Further, to create LBE for cooperative behavior, new structural components will be helpful:

- Players might assume new characteristics or other identities. There could be a sort of *personality construction kit* for building customized identities that players can elaborate between play sessions, as in Internet-hosted MUD (Multi-User Dungeons and Dragons) games.
- Each player could have a different custom *view* of the same LBE. For example, one player might be piloting a futuristic spaceship, while another might be riding a wild animal.
- Players could build and modify games themselves. With sufficiently flexible

simulation engines and a common specification language, LBEs could include tools to build and modify the flow of play.

Games

Trends in game development include new algorithms and a change in emphasis from local games to distributed games.

Algorithms for games are still poor at simulating realistic behavior. Even the most sophisticated adventures rely heavily on prescripted dialogs for character interaction.

Distributed Interactive Simulations (DISs) represent various simulation systems spread over networks which interactively run different scenarios of a specific problem. Before such a large scale of various simulation systems and home platforms sharing a 3D database can occur, DIS protocols must be adopted and severe problems of heterogeneous platform interoperability must be solved, among other things.

18.3 What Are the Multimedia Research Issues?

Many research issues regarding multimedia technologies need to be solved to make any significant progress. Some of them are:

- Random access of optical storage devices still takes too long. Magnetic storage devices are too expensive for the storage of multimedia data. Hence, the research issue is to decrease access time of CD devices, as well as to lower the cost of magnetic storage devices.
- The compression of audio and video sequences, using mode-based techniques, is expected to compress audio and video data at VCR quality to less than 128 kbits/s. Scalability remains a crucial issue.
- The query (and search) for audio and video entries in a database should be content-oriented.

- Communication protocols and networks should support all the different media providing the demanded guarantees. Further, communication and coding methods must allow for interoperability among different systems.

- Operating systems should consider all components which take part in the processing of multimedia data. This should also include distributed system paradigms, such as distributed shared memory and remote procedure call.

- Currently, there is no agreement over the optimal programming interface within a programming environment.

- New metaphors should be explored with respect to user interface audio and video.

- Most multimedia applications have been implemented as improvements on existing technologies. New applications should be researched. An example is an *individual newspaper*, the content and media of which is customized to the individual reader profile.

Many of the above issues require a solid and comprehensive knowledge of multimedia processing. Hence, this book can be used as a working material for learning about current multimedia technologies and for attacking new multimedia research tasks.