Multimedia, Visual Computing and the Information Superhighway

Frederick L. Kitson
Computer Systems Laboratory

HPL-96-18
February, 1996

multimedia, video, graphics, compression, World Wide Web, medical imaging

Multimedia is becoming a mandatory component of most computer interfaces and applications. Medical imaging in particular will be able to exploit these capabilities in concert with data servers that will exist as part of the information superhighway. The ability to connect experts with patients electronically enables care delivery from remote diagnostics to remote surgery. Traditional visual computing tasks such as graphics, volume rendering, computer vision or image processing may also be available to more clinics and researchers as they become "electronically local". Video is the component of multimedia that provides the greatest sense of presence or visual realism yet has been the most difficult to offer digitally due to its high transmission, storage and computation requirements. This paper will address some of the recent innovations in media processing and client/server technology that will facilitate PCs, workstations or even set-top/TV boxes to process and communicate both video and graphics in real-time.

To be published in and presented at the *SPIE Medical Imaging 1996*, Newport Beach, California, February 10-15, 1996.

1 Multimedia Growth

As shown in Figure 1 three major industries that individually generate on the order of a trillion dollars each year are converging. As all forms of media such as film, audio and video become digitized, the communications and processing of such digital data becomes a reality. This in itself presents a major shift for the distributors of such content as well as creating a challenge in ones ability to control and license such lucrative items as a full length movie. The communications industry is anxious to embrace and ultimately capitalize on this trend. As Telco alliances such as Tele-TV which is composed of Nynex, Bell Atlantic and Pacific Telesis Group form to control interoperability so do large cable operators such as TCI and Time Warner. The Tele-TV request for 4.5 million set-tops demonstrates the scale that such devices will be eventually deployed. Computer companies such as HP, SGI, SUN and DEC are also shifting computational products in the direction of video servers and on-line interactive devices. Industries on the boundaries of these titan market segments such as network switching, entertainment software or satellite broadcast are showing startling growth but the real excitement is the intersection of all three industries. Today that intersection is best exemplified by the new applications and capabilities of the Internet. The Internet explosion has often been compared to the California Gold Rush in that it is often not known where the gold nuggets are or how long it might take to mine them but that such a comprehensive broad enterprise must be teaming with wealth. It has also often been pointed out that most of the business generated with the Gold Rush was in the support of such an effort in the form of Levi jeans, Wells Fargo transportation and banking or mining supplies.

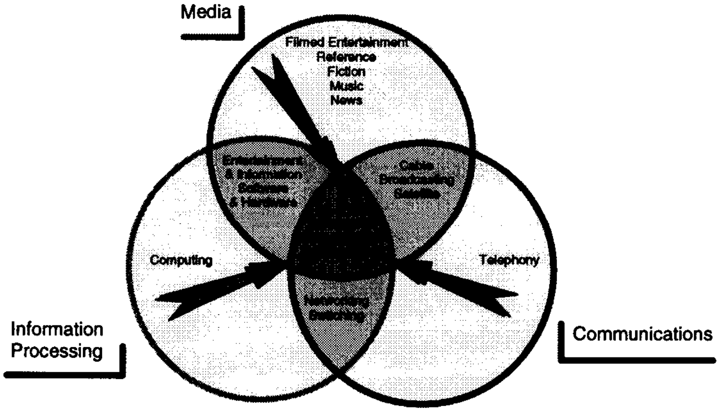

Figure 1 - Growth at the intersection of three industries

The intersection of media and computing is most evident in the plethora of multimedia products from "home PCs", multimedia enabled workstations and the software products in the form of CD based content or newer Web based content. In the medical imaging context, one can view a typical application on a workstation as shown in Figure 2. It is evident that most applications will require and exploit the ability to display 2D images, 1D waveforms, 3D animation, 3D volumetric data and live video. The illustration shown is representative of the system known as NeuroNet[1], at the Center for Clinical Neurophysiology of the University of Pittsburgh Medical Center. "The system is fully integrated, transparently combining the collection, processing, and presentation of real-time data sources, including all physiological monitoring functions with non-real-time functions and extensive on-line database information." As a patient's physiology is dynamic and during surgery, there is extensive monitoring, one can entertain the concept of a closed-loop system. Here measurements such as heart rate can be evaluated against brain-stem stimulation. With video monitoring and appropriate communications, one can utilize collaborative evaluations between teams of neurosurgeons, neuroanesthesiologists and neurophysiologists. As a testimonial to such a system, by stimulating the ear and appropriate monitoring, collateral hearing loss during neurosurgery has all but been eliminated.

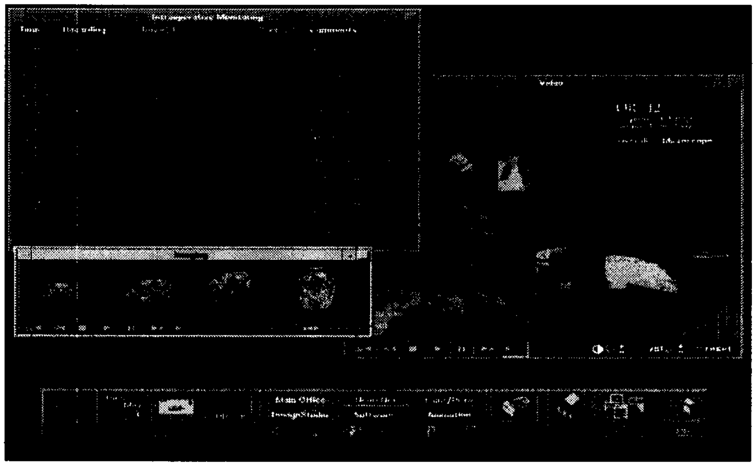

Figure 2 - Medical Multimedia Workstation

2 Visual Computing Technology

In order to support the seamless integration of multimedia, systems must possess the computational components in terms of algorithms and architectures to facilitate the manipulation of data types such as video and graphics. Figure 3 is a layered diagram of the dependencies that may exist. Medical imaging for example might require support for image compression, volume rendering, modeling and high performance communications. This paper will briefly introduce typical results and capabilities that make possible those applications that are inherently visual.

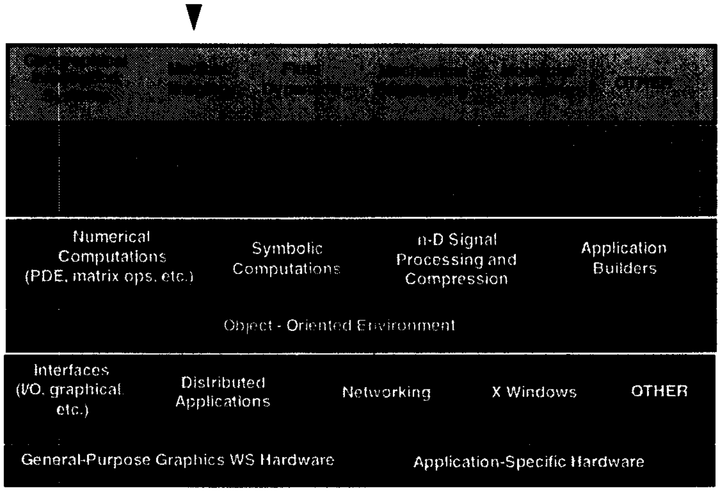

Figure 3 - Tool support for multimedia applications

2.1 Loss-less Compression

In Medical Imaging it is often necessary to store images in a compressed mode to reduce storage and transmission needs yet retain perfect fidelity. A novel algorithm for the loss-less compression^2 of continuous-tone images has been developed which combines the simplicity of Huffman (as opposed to arithmetic) coding with the compression potential of context models, thus "enjoying from the best of both worlds." The algorithm is based on a simple fixed context model, inspired on the insight obtained from the more complex universal context modeling techniques that are also being investigated. The model is tuned for efficient performance in conjunction with a collection of context-conditioned Huffman codes, which is realized with adaptive, symbol-wise, Golomb-Rice codes. While the core of the algorithm is geared toward "smooth," natural photographic images, an embedded run-length coding enhancement, implemented also through Golomb-Rice codes, yields very good compression performance also for "non-smooth" types of images, e.g. text, compound documents, and computer generated

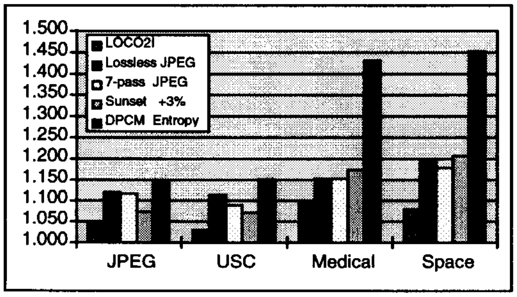

Figure 4 - Typical Loss-less Compression Results

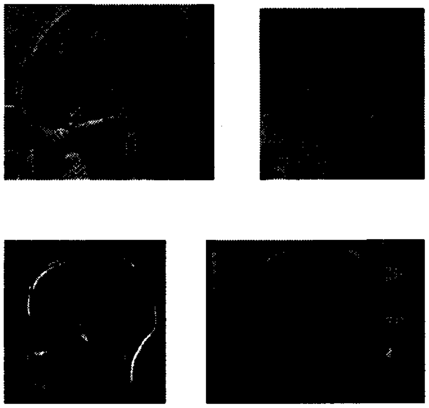

Figure 5 - Example Image Database

graphics. Our algorithm known as LOCO-I (Low Complexity for Images) attains, in one pass, compression ratios similar or superior to those obtained with state-of-the-art schemes based on arithmetic coding, at a fraction of the complexity. In particular, it proves to be superior to the current JPEG loss-less compression system in both compression and complexity. This algorithm has been evaluated on a several standard suites of images (4 shown) with various competitive algorithms with representative examples shown in Figure 5. The results of LOCO-I are shown in Figure 4 with the average compression, normalized to the best known algorithm requiring almost an order of magnitude more running time, on the ordinate. The left-most result is for LOCO-I (where lower is better) indicating its almost universal advantage has been submitted to the ISO JPEG committee[3]. The basic components of the algorithm (coder & modeler) are presented in Figure 6.

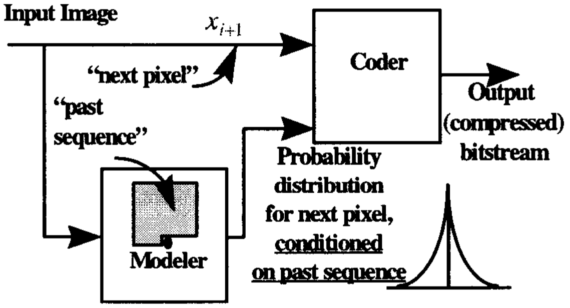

Figure 6 - LOCO Coder Block Diagram

2.2 Volume Rendering

Volume rendering is the process of converting multidimensional data sets into two dimensional images (2D) without an intermediate geometric primitive such as polygons. Often the data is point samples of scalar quantities sampled in three space such as the case with density data in medical imaging. The computation and transmission of 3D data is a further challenge and is relatively quantified in Figure 7. This bar chart shows on a log scale the number of MIPs and MB/sec to process and transmit 3D volume data respectively for Fourier analysis. On the left is the case where 2D images are transformed interactively or at 10 frames/sec. The center case is for 2D data at a resolution of 1024x1024 rendered at a real-time rate of 30 frames/sec. Finally, the rightmost bars illustrate the 3D case at 512x512x512 pixels and full motion (30 f/s). Notice that about 3 orders of magnitude separate the requirements of these three cases. Today's workstations and even the shortly anticipated processors from Intel/HP and others will not provide the necessary raw performance. This is why algorithmic innovations such as Fourier Volume Rendering[4] will be of interest as well as specialized accelerators for this application. Figure 10 shows the spatial data on the left and the 2D transformed image on the right that is its transform. The technique exploits a projection theorem that allows one to take an appropriately filtered slices of a 3D volume transform and perform the 2D inverse instead of the usual 3D data projection. The advantage can typically give one 2 or 3 orders of magnitude reduction in computation for an almost identical image. It also suggests that imaging servers connected by a high-bandwidth interconnect or LAN will be attractive. Another area of interest in post

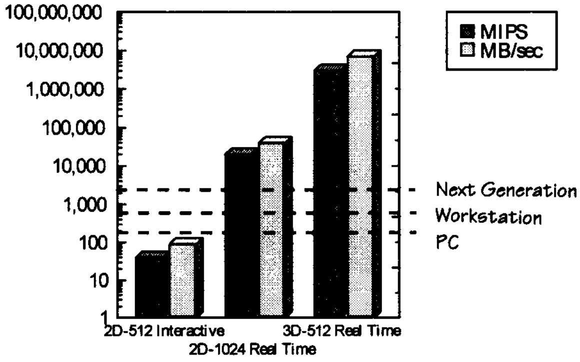

Figure 7 - 3D Computation Comparison

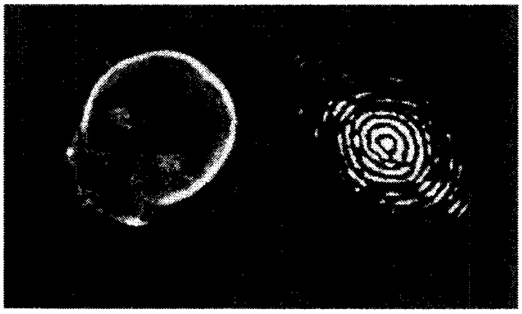

Figure 8 - Fourier Volume Rendering

Figure 9 - Tubular filter

processing algorithms that either segment, enhance or create models of the data for subsequent interpretation. One such process is illustrated in Figure 9 where MRI data with

cross-sections of vessels can be automatically identified after an initialization step to produce a sequence of 2D sections. This data is used to form a continuous 3D tube or vessel model after the "tubular filter" and 3D interpolation is performed. Similar types of processing can then be done to compare this result to stored database models. A configurable custom computing machine, Teramac[5], which is capable of executing synchronous designs of up to one million gates at rates up to one megahertz, has been applied to the problem of artery extraction filtering. Using 7x7x7 filters for artery extraction, it is anticipated one can achieve two orders of magnitude acceleration with half of the processor boards over a competitive single workstation. Again, one sees the need for computation and communications.

2.3 Video Vompression

Much of the focus of work in compression is in supporting video compression standards (MPEG1, MPEG2, H.261 and H.263) on various computing platforms. If one does an analysis of MPEG2 video compression, one would observe that 3.6 Gops (giga-operations) are required for the encoding or estimation of motion vectors out of the approximately 5 Gops total required. Therefore we are investigating various approaches to accelerate the motion prediction problem. One can see the advantage of the approach in Figure 10 where the current frame is on the right and the reference frame in on the left. One encodes the pixels that represent the ping-pong ball by only encoding the offset of the frame in the reference as a first order estimation of the current pixels. Various algorithms are used such as log search, hierarchical search or conjugate direction

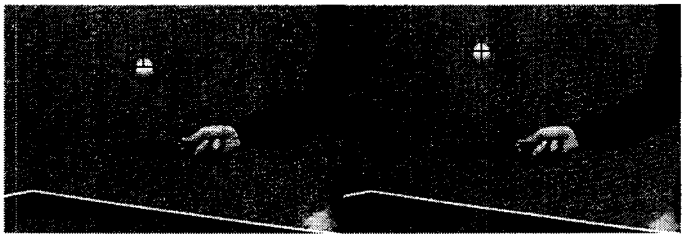

Figure 10 - Motion encoding using previous video frame

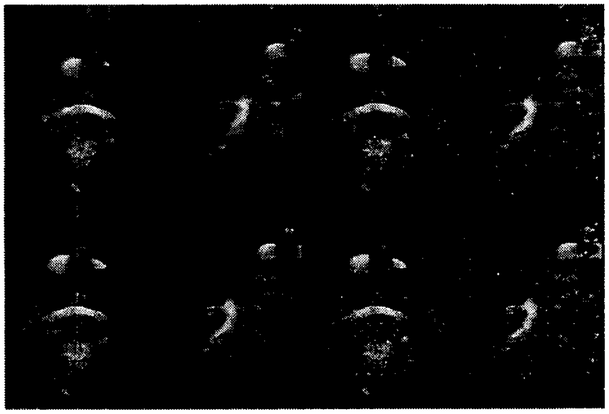

Figure 11 - Reference video frame and frame with motion vectors

search to find the pixels in the reference frame. Of interest are techniques that constrain this search over certain regions to speed up this time consuming process. One such approach[6] using linear prediction of motion vectors based on a maximum absolute difference is shown in Figure 11. By using correlation both in the spatial and temporal domains[7], one can reduce the search area by an order of magnitude with comparable coding fidelity. The overall block diagram is given in Figure 12 which shows how such a technique can be used within the normal MPEG video coding standard. We have also developed a fast algorithm for motion-estimation which yields a 231x speedup over a brute-force motion-estimation strategy exploiting the ability of a computer to do Boolean operations on one-bit thresholded images as opposed to 8 bit arithmetic. Currently, we are actively investigating the use of multimedia instructions for teleconferencing algorithms such as H.263 for a low bit-rate video conferencing standard (8-64 kbps) over the analog telephone line (POTS).

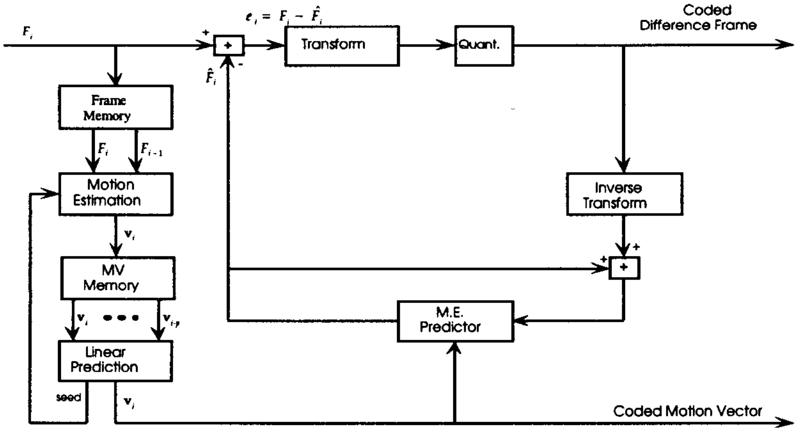

Figure 12 - Linear prediction of motion vectors block diagram

2.4 Video Processing in the Compressed Domain

As applications integrate graphics with video or merely create picture-in-picture functions, there is a need to process video in real-time. This has been prohibitive in the past due to the processing requirements of approximately 1 GOP for video decoding and 5 GOPs for MPEG2 encoding. One example capability is the compositing of several compressed streams into a single MPEG1 or MPEG2 compressed stream so as to enable decoding using a single MPEG1 or MPEG2 decoder. In applications such as multi-party video conferencing, one would like to provide a single composite video stream formed from sub-images of the participants, for example. Often the speaker will be in a larger window than other participants and this will change dynamically. Since video is typically transmitted or stored in the compressed MPEG format, the straight forward or naive approach to achieve this functionality would be to decompress each stream, then scale and possibly decimate the streams to form a summation stream that would then have to be encoded for subsequent transmission. Other applications would require various linear operations on the individual streams. The general problem can then be stated as; can one operate on the compressed data streams directly without the need for the decompression/compression process? The approach is summarized in Figure 13 where a full decompression and subsequent compression are not required. We have had success operating directly on the DCT coefficients in scaling a by a factor of 2, for example, where an algorithm by Natarajan[8] and Bhaskaran gives a low noise result with a computational advantage of a factor of six (5276 ops Vs 880 ops for each output 8x8 block). In general any image or video image with a spatial resolution of NxN that has been compressed using the Discrete Cosine Transform (DCT) coding techniques (e.g. JPEG, MPEG, H.261, H.263), can have the data processed directly in the compressed domain to yield a standards compliant bitstream which when decompressed at the decoder will yield an image or video at a spatial resolution of (N/k) x (N/k) . We have developed exact[9] and approximate scaling algorithms for scale-down by 2 (k=2), scale-down by 3(k=3) and scale-down by 4(k=4).

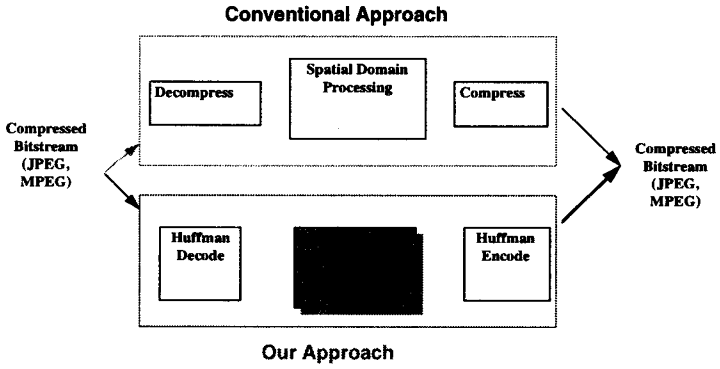

Figure 13 - Operations in the compressed domain

For inverse motion compensation, most MPEG compressed video sequences transmit intra-pictures (I picture) and predictive pictures (P and B pictures) which makes it difficult to edit

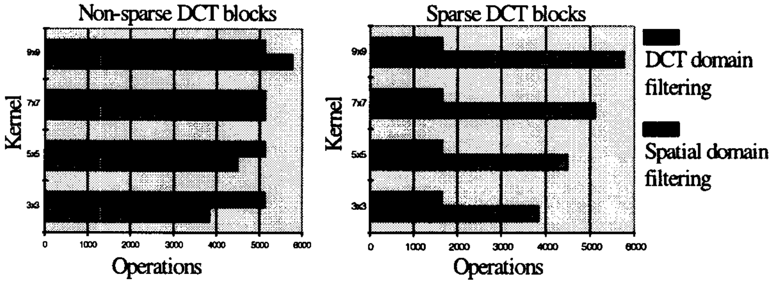

Figure 14 - Operations comparison for filtering in the compressed domain

linear video. We have developed algorithms that can take an IB...BP...P video sequence format in the MPEG domain and convert it to an I....I video sequence. The resulting intra-picture only sequence is now amenable to video-editing (e.g. removing a portion of a video sequence). The spatial-domain approach requires 4320 operations (ops per output 8x8 block) whereas our algorithm requires 2928 ops. For sparse data-sets, our approach requires only 1728 ops.

Another goal is the ability to filter video in the compressed domain. Merhav[10] has developed algorithms that can perform linear filtering by directly manipulating the compressed domain data. This is a very useful image processing operation and our algorithms indicate that for most typical convolution kernel sizes, filtering on the compressed domain requires fewer operations than the conventional approach. Comparisons of the two approaches for typical kernel sizes are shown in Figure 14.

In many video special effects generation, there is a need to merge two sequences. We have examined a special case of the merge operation where in given a compressed stream, we would like to cut out a small region of a frame from this stream and insert another stream in its place, (e.g. we might want to combine a video sequence with the station logo). This is essentially a pixel-wise multiplication in the spatial-domain. We have developed algorithms[11] that perform this operation in the compressed domain. There is a 47% reduction in the operations needed to perform this task using the compressed domain approach versus the spatial-domain approach.

In some applications, the viewing experience can be personalized to the user's requirements. For instance, during the viewing of a football game, the user might want to focus on a single player's

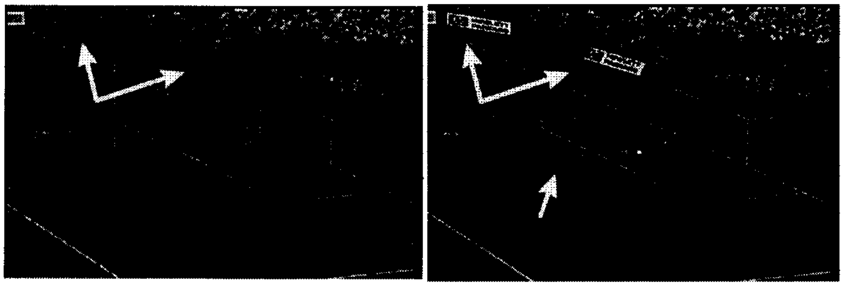

Figure 15 - MPEG2 frame with image merge & graphic

actions. This user driven focus might be accomplished by the user first selecting a region of interest on the screen and then the processor would track the object within this region and perhaps display the object within the scene using say, a lighter background for the object. New applications such as advertisement insertion or overlay in video streams requires the mapping of 2D images, textures or 3D graphics projections into live video sequences. Figure 15 shows an example using MPEG2 resolution images from a 1994 World Cup Soccer match. The area with the "Coca Cola" billboard has been identified for example in the compressed domain and the area is then tracked by its motion vectors. The tracked area is then replaced with the appropriately transformed texture "Hewlett Packard" in this case forming the "Hewlett Packard" billboard on the right. Other applications might require the tracking of objects such as the soccer ball whereby a synthesized trailer might be color coded to indicate the objects velocity.

2.5 Video Content Retrieval

Content-based image retrieval systems[12] such as the QBIC[13] (Query by Image and Video Content) at IBM Almaden Research Center or Chabot[14] which is a relational database system for images are reaching a level of utility that shows such systems will serve to enhance medical imaging applications. QBIC, for example, supports queries of large image and video databases based on example images, user-constructed sketches and drawings and selected color and texture patterns. In Chabot, a database management system (DBMS) that allows one to include image analysis and information-retrieval tools in the query process. Searches can combine text and image operators such as identifying an object general shape or location. Figure 16 illustrates this with a search initiated by the lower diagram that is looking for a video frame with an oval shape in the general location indicated. Many applications are looking for scene transitions to aid in editing.

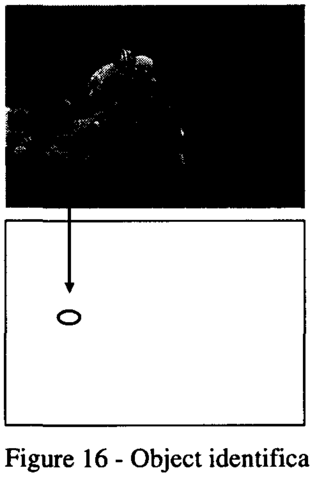

Figure 16 - Object identification in video

2.6 Audio Compression

Ironically, the audio decompression function which when implemented in floating point consumes about 30% of the cycles compared to video decompression even though audio constitutes only 10% of the data. We have been investigating fast integer-based implementations for MPEG layer-1 and layer-2 audio decoding. By careful algorithmic enhancements[15] (e.g. use of a fast DCT method and judicious use of embedding some of the multiplies of the DCT process in a post-process) it has been found that the computational complexity can be reduced. Further reductions in computation complexity have been obtained by judicious choice of audio data truncation in the sub-band domain and truncation of the weighting coefficients. The truncation effects were found to be subjectively not objectionable.

Of greater interest is the reality of audio over the Web via such technology as RealAudio (http://www.RealAudio.com) for modems and Streamworks for ISDN. A RealAudio server can deliver AM Radio quality broadcasts by front-loading an audio file for latency and then the decompression is done on the fly at the user's PC.

2.7 Graphics Processing

When using a video sequence such as in Figure 15, one may want to identify or track objects or regions such as soccer players. In Figure 17, we show that a tracking mode may be turned on with an identified player. A dynamic hot spot on the screen is then established so that if one were to click on the region of the screen associated with the player at that particular time, then the application can execute a link such as shown, so that additional information on the player might be displayed for example. We call this notion HyperVideo.

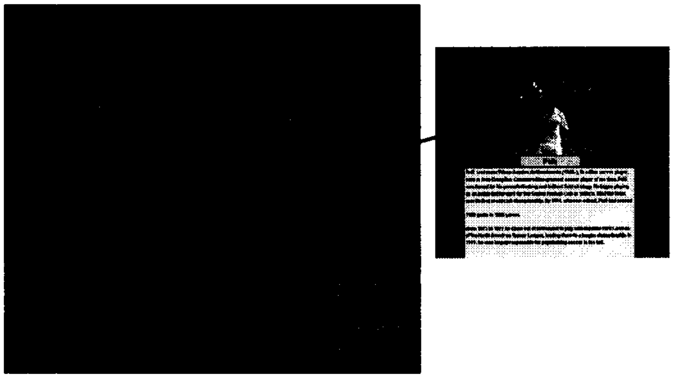

Figure 17 - Graphics in video & dynamic "hot spots"

Figure 18 - Image composed of video, 3D graphics and 2D images

Graphics will be required to present user interface elements for navigation and presentation as in a Web application or interactive games. An application may be downloaded into ones PC or even a set-top box (to be discussed later). Performance requirements of realistic graphics will mandate that close to one million anti-aliased vectors and texture mapped polygons per second be integrated with video processing. This will enable animated and colorful gripping interfaces with a mixture of 3D animation, video and images as shown in Figure 18. Three-dimensional graphics will support advanced navigational systems, games and new applications such as home shopping. A typical client for graphics and video will have an architecture with the components as in Figure 19. The key blocks for multimedia such as image, graphics and video can be merged into a "media processor" which is discussed in the next section. One opportunity that the new digital communications infrastructure will support will client-server graphics as well as video and audio. One can partition the application so that a large database may reside on a server (see Figure 20) with only the portion of the visible model necessary for the client or home terminal to be transmitted and processed. We are addressing two algorithmic areas. The first deals with preprocessing and spatially sub-dividing graphics data bases for visibility and then transmitting only the potentially visible sub-set of the data base from a server to the client for rendering. This provides savings in memory, network bandwidth and compute requirements. The second

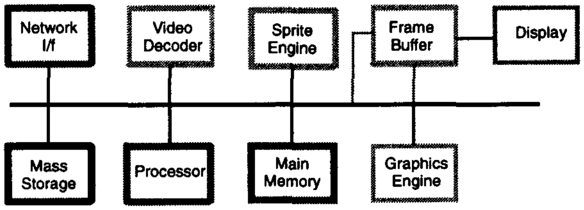

Figure 19 - Basic system components

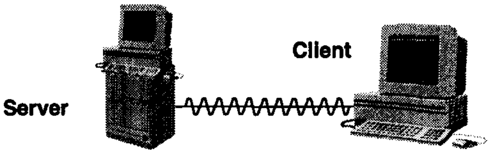

Figure 20 - Client-server graphics & video

algorithm deals with multi-resolution representation for graphical databases.

Video may be fed from the server and the graphics composited on the workstation, PC or set-top with some of the innovations discussed above. Microsoft refers to the next generation of home clients as "communicating PCs" which have the aforementioned video and graphics capability with a broadband interconnection. Often, ATM (Asynchronous Transfer Mode) is mentioned as a preferred packet-oriented protocol for multimedia traffic because of its notion of QOS (Quality of Service). Although the PC represents the present information appliance (and game platform), there is still the concept of a multimedia communications device attached to the TV set, collectively known as an interactive TV system. The three basic components of an interactive TV architecture are content (information servers), a network, and the set-top box [16].

2.8 Physically Based & Constraint Modeling

Another technology area that holds promise for the design of multimedia interfaces is physically based modeling[17] and constraint based modeling.[18] The concept is to assign relationships between objects such as "next to" or "on top of" as opposed to precise geometric specifications such as coordinates. By modeling physical attributes such as friction and gravity, one can also control predicted behavior. As shown in Figure 21, a robot arm can be modeled by forces and behavior and the actual grasping and movement of the block to the conveyer belt is done automatically through a constraint solver. This greatly simplifies the generation of an animation or physical simulation.

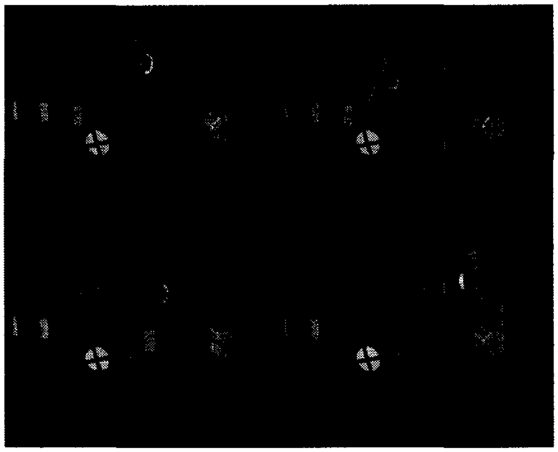

Figure 21 - Constraint specified robot arm

2.9 Central Processor Unit and Media Processors

A central processor unit and associated memory will provide basic control functions in the set-top box (Figure 22). The integration of video and graphics processing with the CPU would form

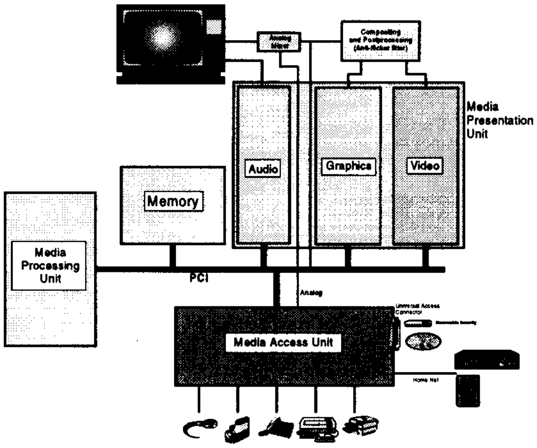

Figure 22 - Set-top block diagram and peripherals

a second or third generation "Media Processor". Such a processor will have a specialized capability to support MPEG2 video decompression, audio decompression and some level of graphics capability. Strategies such as those indicated in the design of VLIW processor design may be enhanced with specialized instructions or co-processors. Memory architecture and the algorithms that utilize the minimum dynamic memory will have a big impact on the viability of such systems for consumer deployment. Again client-server technology will play a strategic role. The choice of memory and CPU will be constrained by the low price points that a set-top box will probably sell in the $300-$700 range. An intelligent set-top box provides a great opportunity to introduce a number of peripherals and services into the home. A set-top box will provide a connection and protocols for such peripherals. A printer connected to a set-top could augment home shopping by printing coupons, invoices and copy of orders that a user places. A CD-ROM could supplement off the air programming by mixing information from the CD-ROM with information on the cable. DVD which is a 7 fold increase in the density of today's CD offers the opportunity to distribute higher quality video, multiple audio tracks and executable code for PC applications. This consumer oriented technology that is supported by many of the house-hold names will be introduced this year. Integration with existing voice telephone also provides some interesting possibilities. Other functions such as image processing and telephony will be supported as required by applications noted above but are not discussed further here.

3 Information Superhighway

The birthplace of Web is CERN (European Laboratory for Particle Physics) where "World Wide Web" or WWW is considered to be the "universe of network-accessible information, an embodiment of human knowledge"[19]. The Web is also regarded an a gigantic on-line multimedia encyclopedia due to its recent support for dynamic graphics objects, images, audio, video that are hyper-linked. The key feature of the Web is that multimedia objects whether text or video are linked by a vast collection of HTML (hypertext markup language) over the Internet. The

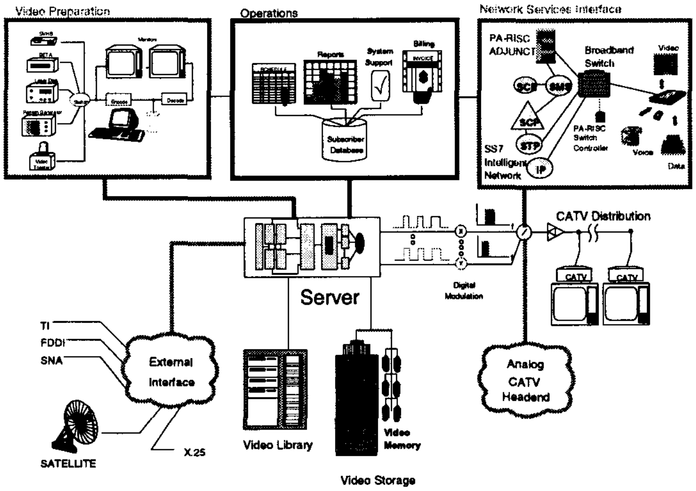

Figure 23 - Video-on-Demand system

consistent method for specifying all Internet resources so that Web browsers can locate or address a server is URLs (Uniform Resource Locators).

Such a system of interconnected heterogeneous data servers represents the opportunity and challenge in terms of multimedia access. With such limited bandwidth and almost infinite data, it has engendered a scramble by network and cable providers to offer the connectivity of the web with the bandwidth of cable or the switching functionality of the Telcos. The realization of a distributed multimedia system the connects every home and business is the vision of the "information super-highway".

3.1 Cable Modems

Although there has been and continue to be substantial investments, trials and product announcements, there have been some setbacks to the realization of a true digital video-on-demand (VOD) system depicted in Figure 23. Scaling video servers to the level required for full deployment in the commercial market is still years away. The $300 fully interactive set-top is also under active development but requires state-of-the-art multimedia and communications and it may take a few years to achieve this cost goal. Video servers which have the general design architecture as shown in Figure 24.

Figure 24 - Video server architecture

Therefore, the next real advance in the information superhighway will be the deployment of hybrid-fiber coaxial system as shown in Figure 25. Digital data will be linked to the distribution hub where five fibers will support each community hub. This arrangement allows the management of about 2000 homes. Each home will have a cable modem that is attached to ones PC or other information client such as a set-top. The major advantage is that tens of megabits per second will be available for delivery of data or even MPEG2 video to the home as opposed to today's tens of kilobits per second that limits the applications that use images, audio or video. By using digital QAM modulation, a 30 Mbit/sec data stream can exist in a 6 MHz analog channel. Also there is a "back channel" for interactive applications that is a shared communications channel in the units of megabits per second. We have developed appropriate communications protocols for such a system and are actively participating in the cable data consortium which includes Intel, AT&T, Hybrid Network as well as standards bodies to define international cable data standards. The real advantage of such a system is its ability to operate with the inherent latency and the ability to load balance the multiple upstream channels. Another advantage is that such channels can be added incrementally to an existing cable infrastructure with existing analog cable channels.

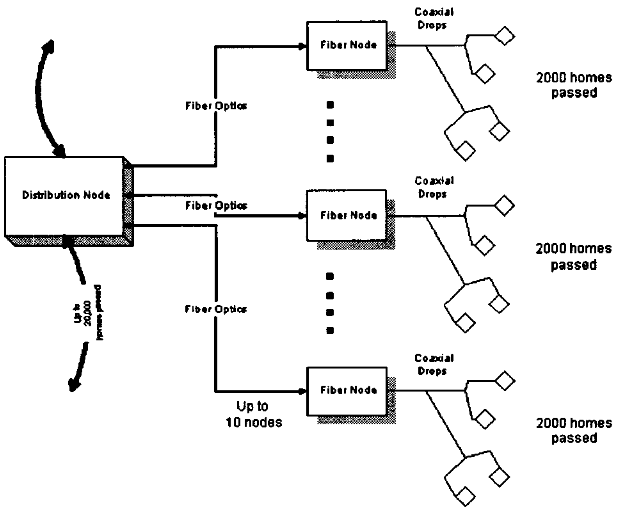

Figure 25 - Hybrid Fiber Coax Distribution

3.2 3D Web Interfaces

The Web has spawned many innovations that are quickly becoming integrated with browsers through mechanisms such as SUN's JAVA, Microsoft's Blackbird, Macromedia's Director or VRML[20]. Java for example allows one to download a platform independent code that enables the execution of code locally. Java[21] is built on C and C++ languages optimized for client-server on-line applications. Code is dynamically linked on the fly through an interpreter so that an application runs without pre-compilation. A Java application or applet can be sent from the server to the client running HotJava which is the player that automatically runs the application. The server can then continue to update the application. Such capability is now built into NetScape for example and allows one to run CD ROM code and a remote multimedia application through the use of its "socket" technology. Blackbird is a Visual BASIC style authoring environment. Like MS Word support for HTML, one will have HTML to Blackbird conversion. The Blackbird server acts as an information manager[22] giving the client side memory and routes data to an application that has already been given a template and an "OLE control container" which manipulates the data. Virtual Reality Modeling Language or VRML is also enabling graphics applications that exploit client server capabilities. An example of a 3D interface metaphor is illustrated in Figure 26 as a 3D house. One can navigate into the "garage" to go on a "trip" to Hawaii or virtual 3D mall or even get mechanical advice. Move into the study for Video-on-Demand, outdoor weather viewing or pull a 3D book out of the library and use the Web to enter Cyberspace. One such prototype of the nearly 2000 expected in 1996, is Virtual Soma (South of Market Street in San Francisco, http://www.hyperion.com/planet9/vrsoma.htm - Intervista Software) which consists of three blocks today, complete with photo-realistic facades. One can enter buildings which is also possible without such an interface but would require that you know each URL. VRML 2.0 will allow one to customize their avatar or the personal appearance of ones virtual agent used for multi-user encounters.

Figure 26 - 3D information interface

4 Medical Imaging Future Applications

The ultimate application of communications and visual computing may be what is facetiously referred to as the "long-distance operators"[23]. The Pentagon has been funding the field of remote medicine for tele-surgery under the name of MEDFAST or Medical Forward-Area Surgical Telepresence. The concept is to place the surgical unit in areas of conflict, accidents or natural disasters and enable a surgeon to control through force feedback devices such as surgical instruments the "touch" of a remote operation as suggested by Figure 27. ARPA is also funding the development of a vehicle to transport such a system known as M3 for Mobile Medical Mentoring. The entire system could use superhighway information access or specialized aircraft and satellite systems. As a testimonial to the feasibility of this concept, researchers have used a prototype system to operate on a pig at the University of California at San Francisco, with assistance by consultants from SRI's electronics technology laboratory.

Figure 27 - Robotics manipulator under surgeon's control (PS)

Finally, as a visual example of the opportunity that wireless video and data communications offer, there is a picture of Sharp Electronics vision of a "multimedia Swiss-Army knife" in Figure 28. Many portable devices are under development which bring wireless video integrated with graphics to potentially provide the medical imaging appliances of the next century. Again, technologies such as client server-graphics and model based compression such as MPEG4 will help to engender these portable clients. Although such devices will play a role in the hospital, a greater opportunity exists in support of home care with companion home imaging and instrumentation.

Figure 28 - Portable Multimedia (Sharp Electronics)

5 Acknowledgments

Most of the work reported has been developed by members of the Visual Computing Department at HP Labs in Palo Alto. In particular, the video work was done by Vasudev Bhaskaran, Balas Natarajan, Robbie Armitano (Georgia Tech) and Neri Merhav (Technion - Israel Science Center); the graphics work by Deven Kalra, Tom Malzbender, Michael Goss, Susan Spach and Kevin Wu; the loss-less compression work by Gadiel Seroussi and Marcelo Weinberger; the audio work by Mat Hans (Georgia Tech); the cable modem by Ran-Fun Chiu and Mehrban Jam.

6 References

[1] Sclabassi, et al., "NeuroNet: Collaborative Intraoperative Guidance and Control", IEEE Computer Graphics and Applications, January 1996, pp39-45.

[2] Weinberger, G. Seroussi, G. Sapiro, "LOCO-I: A Low Complexity, Context-Based, Lossless Image Compression Algorithm," HP Labs Technical Report #HPL-95-62, June 1995.

[3] LOCO-I: new developments, Committee ISO/IEC JTC 1/SC 29/WG 1 (ITU-T SG8), Document N 245, Oct. 31 1995.

[4] Malzbender, Fourier "Fourier Volume Rendering", ACM Transactions on Graphics, Vol. 12, No. 3, July 1993, pp. 233-250.

[5] Amerson, R. Carter, W. Culbertson, P. Kuekes, G. Snider, "Teramac - Configurable Custom Computing", IEEE Symposium on FPGAs for Custom Computing Machines, April 1995.

[6] Armitano, R. Schafer, F. Kitson, V. Bhaskaran, "Linear Predictive Coding of Motion Vectors", Proceedings of the Internal Picture Coding Symposium, pp. 223-236, September 1994.

[7] Armitano, R. Schafer, F. Kitson, V. Bhaskaran, "Motion Vector Estimation Using Spatio-Temporal Predeiction and Its Application to Video Coding", SPIE Conference, Digital Video Compression: Algorithms and Technologies, San Jose, Jan 27 - Feb 2, 1996.

[8] Balas Natarajan, Bhaskaran Vasudev, A Fast Aproximate Algorithm For Scaling Down Images In The DCT Domain, IEEE International Conference on Image Processing', Oct. 1995.

[9] Neri Merhav, Bhaskaran Vasudev, A Fast Algorithm For DCT-Domain Image Downsampling, Accepted for presentation at ICASSP'96, Atlanta.

[10]Neri Merhav, Bhaskaran Vasudev, A Fast Algorithm For DCT-Domain Filtering, HPL Tech. Report, HPL-95-56.

[11]Neri Merhav, Bhaskaran Vasudev, A Fast Algorithm For DCT-Domain Masking, HPL Tech. Report, HPL-95-104.

[12], Raghavan, V.V., "Content-Based Image Retriveval Systems", Gudivada ,IEEE Computer Magazine, September 1995, pp. 18-31.

[13]Flickner et at, "Query by Image and Video Content: The QBIC System", IEEE Computer Magazine, September 1995, pp 23-32.

[14]Ogle, M. Stonebraker, "Chabot: Retrieval from a Relational Database of Images", IEEE Computer Magazine, September 1995, pp 40-48.

[15]Hans, V. Bhaskaran, "A Fast Integer Implementation of an MPEG-1 Audio Decoder", HP Laboratories Technical Report, HPL-96-03, January 1996.

[16]Furht, D. Kalra, F. Kitson, A. Rodriquez, W. Wall, "Design Issues for Interactive Television Systems", IEEE Computer Magazine, May 1995, pp. 25-38.

[17]Kalra and A. H. Barr., "Modeling with time and events for computer animation", Eurographics Proceedings, Cambridge, England, 1992.

[18]Kalra and A. H. Barr."A unified framework for constraint-based modeling",Visual Computing -- Integrating Computer Graphics with Computer Vision, Proceedings of Computer Graphics International, 1992, Tokyo, pp 675--696, June 1992.

[19]Web Publishing with Microsoft Word, Herb Tyson, SAMS.net Publishing, Indianapolis, IN, 1995.

[20]Pesce, VRML: Browsing & Building Cyberspace, New Riders Publishing, Indianapolis, IN, 1995.

[21]Ritchey, JAVA, New Riders Publishing, Indianapolis, IN, 1995.

[22]Amdur, “The Scene Is Set For Multimedia On the Web”, NewMedia, Nov. 1995, pp. 42-52.

[23]Gourley, “Long-Distance Operators”, Popular Science, September 1995, pp. 64-67.