# Interactive Video from Desktops to Settops

Frederick Kitson, Vasudev Bhaskaran,
Deven Kalra
Computer Systems Laboratory
HPL-95-58
June, 1995

video, compression, MPEG, settops, desktops, graphics

Video is the component of multimedia that provides the most visual realism while placing the most stress on a computer system. The capture, processing, transmission, digital storage and display of video requires a delicate balance between available MIPs, MB/sec and MB of dynamic and static memory. Multimedia applications of the future will combine interactive video and graphics in new and exciting forms. This paper will also address some of the issues and innovations in engendering *media enabled* computer systems from conventional desktops such as workstations with modern RISC processors to the next generation of consumer computers known as "settops". To provide a specific example, a MPEG1 decoder that is capable of real-time playback of video and audio on HP's RISC-based workstations will be described. The desktop community is seeking "TV-like" functions such as surround sound and broadcast quality video while the home consumer desires "computer-like" interactivity and connectivity.

Internal Accession Date Only
To be presented at and published in the proceedings of the *NEC Symposium in Multimedia,* Tokyo, Japan, June 7, 1995.

## 1 Introduction

Multimedia is often defined in terms of data types such as video, audio, images, graphics, numbers and text. Today tools exist to manipulate and access text, for example, via word processors or numbers via spreadsheets. With the capabilities of current Digital Signal Processors, audio has also become a supported data type. In the next few years, full-motion video will achieve this status. To this end, computer designers are commissioned to architect the fundamental capabilities to capture, manipulate, store and transmit video. Because of the performance requirements necessary to achieve such *dexterity*, compression algorithms are presently mandatory. Compression and decompression support have therefore become enabling technology for multimedia systems. Fortunately some standards have gained popular support such as MPEG for motion video so that VLSI manufacturers and application developers can create interoperable systems.

As with audio, the first systems to support video have been realized with specialized processors that are tuned for DSP or video in particular. As general purpose processors achieve higher MIP ratings and adapt to this new media type, video processing will come under the domain of workstation applications. This paper will present an example of both situations. First a summary of HP's Precision Architecture (PA) RISC processor will be presented to show how a high performance general processor can support an efficient system for handling compressed video and audio. MPEG1 playback was chosen as a key goal since it is a good match in terms of complexity and current applications such as CD-ROM support.

This work can be extended to other operations on video such as scaling and merging of video streams for teleconferencing. New applications such as medical imaging can be entertained with support for 2D and 3D data at video or interactive rates (10-30 frames/sec.).

Consumer video systems will be a catalyst for the development of an aggressive price/performance point with the primary objective of video decompression in real time and general purpose multimedia support as a secondary requirement. This low cost interactive video possibility is engendered by the confluence of video compression, digital processors, memory integration and communications processing with cable TV infrastructure. The *settop* will provide the interface from the communications interface or cable to the video monitor or television. This interactive processing device will demodulate, decode, decrypt and decompress digital video and audio streams as well as process analog video. It will be the customer interface for viewing and service selection for such applications as movies-on-demand, music-on-demand, games-on-demand and home shopping. HP is creating a *consumers computer* that represents a "Trojan Horse" into the home for information access. It will contain high volume, low-cost media processors and interfaces that can span from settops to desktops. We are meeting this challenge with innovative algorithms and architectures to meet the high performance requirements and low cost. The back channel for interactivity is key to enhanced applications and services for the next generation settop. Video and multimedia servers will support settops through a client-server relationship.

## 2 MPEG1 Decompression on HP Workstations

The decoding of a MPEG1 bitstream as performed on HP's workstations[1] includes (a) system level decoding to extract the timing information and demultiplexing of the compressed video and audio streams, and (b) video decoding[2] to decompress the MPEG1 video data. Audio decoding is done as well but this aspect will not be covered here. Figure 1 shows a block diagram of the MPEG1 video decoder.

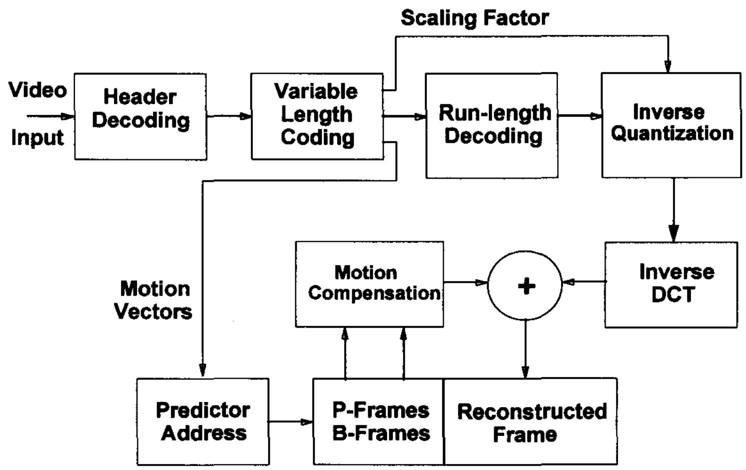


Figure 1 : MPEG1 Video Decoder

Video decoding consists of these steps:

1. Video sequence header decoding to extract parameters of the video sequence such as picture rate, bit rate, image size, etc. For each group of pictures (GOP), identify picture type, e.g. I, P or B picture. For each picture and for each slice within the picture, determine the quantizer scale.
2. For each slice, decode each macroblock. Macroblock layer decoding consists of extracting the motion-vectors from the coded stream and then extracting the DCT information for the blocks within the macroblock.
3. The DCT information is huffman coded. Thus huffman decoding is performed to decode the variable-length codes into fixed-length symbols.
4. Inverse quantization is performed on the huffman decoded data.
5. For each 8x8 block of the inverse quantized data, a 8x8 inverse DCT is computed. This transforms the data back to the image domain.
6. Motion-compensation is then performed if needed. For P blocks and B blocks, motion-compensation consists of taking the inverse DCT output and adding it to the reference block(s) pixel values; reference block address is given by the motion-vector information decoded at the macroblock layer.
7. Finally, the image domain data is displayed. The display step includes color conversion from the YCbCr color space to the RGB space. Since Cb and Cr pixel data is half the resolution of

the Y data, upsampling needs to be performed during or prior to the YCbCr to RGB conversion phase. Additional upsampling of the pixel data may be required for display, e.g. the player might have to display the image in a larger window than its original resolution.

Steps 2-7 are compute-intensive and are the main bottlenecks to real-time MPEG1 video playback for a software based video player. In a practical implementation, some form of error-concealment must also be employed during video decoding. In the next section, we describe some of the optimizations incorporated in HP's MPEG1 video player.

## 3 Algorithm and Architectureal Enhancements

### 3.1 Enhancement Methodology

The basic approach was to examine the workload associated with each step of the decoding process outlined in the previous section and then develop algorithms for some of these steps that would lead to a reduced workload. The performance goal was to get a 10 - 15 fps playback of SIF resolution (352 x 240) MPEG1 compressed video and audio assuming that all of the enhancements were restricted to the algorithm level only.

A simple analysis of the video decoding steps outlined in the previous section indicated that the bulk of the execution time was spent in the IDCT step (46.4%) followed by the Display step and then the Motion-compensation step. Other steps in the decoding process consumed negligible time. Thus, algorithm and architectural enhancements were primarily targeted at these steps of the video decoding process.

### 3.2 Video Decompression - IDCT Optimization

In MPEG1 compressed video, an analysis was performed on the bitstreams. It was observed from this analysis that the IDCT computations were often performed on sparse matrices. Thus, if one could determine the nature of this sparseness, one could reduce the computation load of the IDCT. In order to determine the sparseness without additional overhead, it was found that by viewing the huffman decoder, inverse-quantization and the IDCT computations as a single system, it is possible to develop a computation procedure that reduced the workload for these three steps combined. This is the approach that is adopted in HP's MPEG1 player.

Inverse quantization can be performed within the huffman decoder, thereby, reducing accessing the same data twice. A low complexity IDCT algorithm was developed; its worst case performance is 80 multiplies and 464 additions for a 8x8 block. By exploiting the sparseness information, this IDCT algorithm yields an average performance of 46 multiplies and 253 additions for a 8x8 block. A lookup table based approach can be used for the multiply operation since the constants used in the IDCT were relatively few. Lookup table accesses are memory accesses which may be time-consuming. Instead, in the IDCT, the constants were chosen such that the multiply operation can be performed with a minimum number of shift-and-add operations and yet maintain good accuracy within the IDCT. The shift-and-add operation was further restricted to shift by 1, 2 or 3 bits since these operations are native instructions for the PA-RISC CPU.

With algorithmic enhancements only, the video decompression tasks breakdown on a PA-RISC CPU is as shown in Table 1.

Table 1 : MPEG1 Video Decompression Tasks Relative Execution Time

| | |
|---|---|
| Header Decode | 0.1 |
| Huffman Decode | 7.5 |
| Inverse Quantize | 2.4 |
| IDCT | 38.7 |
| Motion Compensation | 18.3 |
| Display | 33.0 |

Note that the IDCT, motion-compensation and display tasks are still the dominant tasks. Architectural enhancements were then explored to speedup these tasks.

### 3.3 Video Decompression - CPU Related Architectural Enhancements

In terms of architectural optimizations, several PA-RISC *multimedia* instructions[3] were added. These instructions allowed parallel operations of several simple arithmetic operations by operating on subword data in the standard 32 bit integer datapath. For instance, the 32 bit integer ALU was partitioned so that it could execute a pair of 16 bit arithmetic operations in a single cycle with a single instruction. Arithmetic operations that were accelerated using this strategy include add, subtract, average, shift-left-and-add and shift-right-and add. These operations also integrated several functions within the parallel operation so as to yield a very efficient instruction as illustrated in the following example.

Consider the PARISC multimedia instruction *HADD,ss,ra,rb,rc* (this instruction performs addition of the two 16 bit quantities in registers ra and rb and saturates the results so that it does not exceed a preset maximum and minimum value. The saturated 16 bit results are then deposited into the 32 bit register rc. Without this multimedia instruction 10 operations have to be performed to get the desired 16 bit results. The multimedia instruction on the other hand, yields the two signed saturated 16 bit results in 1 cycle.

Note that the add, subtract, shift-left-and-add and shift-right-and-add are used intensively within the IDCT and thus led to additional speedup of the IDCT task due to architectural enhancements compared with the algorithmic enhancements performed on the IDCT as described earlier. The motion compensation task is not amenable to any algorithmic enhancements. In this case, the average instruction as implemented in the architecture was extensively used so that for a B block in MPEG1, two averaged pixels can be computed in a single cycle. Without the multimedia instruction, this operation would require four cycles.

In the PA7100LC PA-RISC CPU, approximately 0.2% of the silicon area was added to provide these multimedia instructions. There was no impact on the processor's cycle time and furthermore, the area used was mostly empty space around the ALU; thus one can claim that the multimedia instructions has contributed to more efficient area utilization. The PA7100LC has dual integer ALUs; thus, for the 16-bit multimedia instructions, a conservative speedup of four is obtained for 16 bit operations compared with the conventional 32 bit ALU.

### 3.4 Video Decompression - Graphics Subsystem Architectural Enhancements

The CPU load for the display step was significantly reduced. The strategy here was to exploit the capabilities of the graphics subsystem within the HP workstation. The graphics subsystem is capable of handling YCbCr data and can perform the upsampling of the Cb and Cr data and perform conversion from YCbCr to RGB. Furthermore, to reduce frame buffer requirements, the HP workstation's graphics subsystem architecture is such that 24 bit pixels can be kept in a dithered 8 bit mode. Color compression[4] allows use of 8 bit frame buffers in low-cost HP workstations. The dithering is done in a dynamic manner within the graphics subsystem and leads to very good quality rendering of the original 24 bit RGB data. The graphics subsystem is also capable of scaling the video during display; this permits displaying a SIF resolution video at twice its size without increasing the bus traffic from the CPU to the graphics subsystem and without increasing the framebuffer size. This leveraging of low-level pixel manipulations close to the frame buffer between the graphics and video streams contributed significantly to realizing real-time MPEG1 decompression.

### 3.5 Performance

Algorithmic and architectural enhancements as well as leveraging of functions within the graphics subsystem in HP's workstations yields real-time playback of MPEG1 compressed video and audio streams. For a typical video clip at SIF resolution and 30 fps, the maximum MPEG1 decode rate is 33.10 fps for the HP 712 workstation with an 80 Mhz processor. Here, the decompression is for the video only (the audio is not decompressed). The HP 712 workstation incorporates the *multimedia instructions* described in section 3.3.

The performance of this MPEG1 player is compared against performance figures that have been reported elsewhere for MPEG1 players on other computing platforms. This comparison is shown in Figure 2 (the performance figures are for video decode only). Note that one of the recently announced MPEG1 players for the Pentium has achieved 20-25 fps on a 90MHz Pentium.

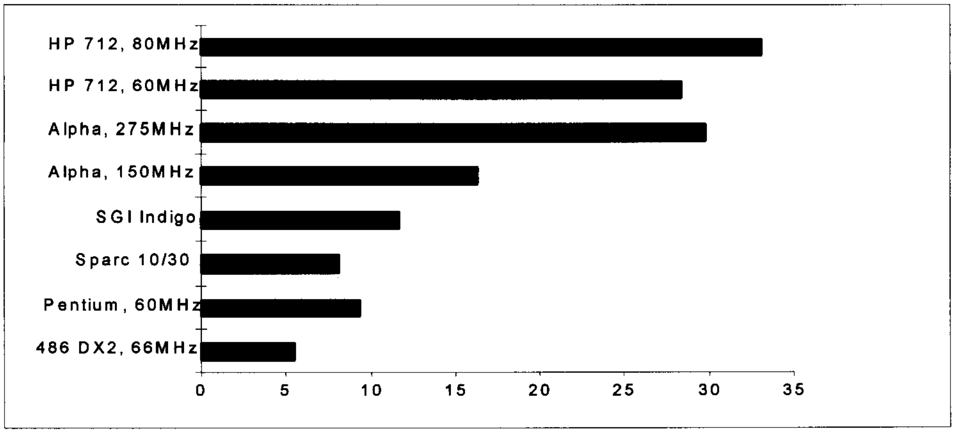


Figure 2 - MPEG1 playback framerate on various platforms (based on data reported elsewhere).

In Table 2, we show the performance comparisons between an unenhanced MPEG1 video player and the HP MPEG1 video player. This table illustrates the performance that can be obtained by enhancing the MPEG1 decompression process at the algorithm as well as the architecture level.

Table 2 : MPEG1 Video Performance On Unenhanced and Enhanced Systems

| | |
|---|---|
| Unenhanced, 720, 50MHz | 9.7fps |
| Algorithm enhancements only - software level, 720, 50MHz | 10.9fps |
| Architecture enhancements only, 712, 60MHz | 11.1fps |
| Algorithm and Architecture Enhancements, 712, 60MHz | 26.2fps |

### 3.6 MPEG1 Summary

A high-performance MPEG1 video player has been developed for HP's PA-RISC workstations. The video player attains real-time playback through synergistic algorithm and architectural enhancements. The algorithm enhancements are applicable to any general purpose CPU. The architectural enhancements have negligible impact on the silicon area; however, they yield significant performance gains. The MPEG1 core as enhanced in this work is similar to the JPEG core, H.261 core and the MPEG2 core. Thus, the enhancements reported here should improve the performance of JPEG, H.261 and MPEG2 decompression on HP's multimedia-enabled PA-RISC workstations. Through higher levels of parallelism, the methodology adopted here would lead to real-time MPEG2 playback at CCIR601 resolution as CPU technology improves over the next few years. Table 3 adopted from Konstantinides and Bhaskaran[5] gives approximate MIPS requirements for encoding and decoding MPEG2 which will be the broadcast standard covered in the next section on settops. Note the dominance of Motion Estimation for encoding.

Table 3 : MIPS Requirements for MPEG2 (Konstantinides, Bhaskaran)

| COMPRESSION | MIPS |
|---|---|
| RGB To YCrCb | 108 |
| Motion Estimation | 3648 |
| (i.e. 25 searches in a 16x16 region) | |
| Coding Mode | 160 |
| Loop Filtering | 0 |
| Pixel Prediction | 108 |
| 2-D DCT | 240 |
| Quantization, Zig-zag scanning | 176 |
| Entropy Coding | 68 |
| Reconstruct Previous Frame | |
| (a) Inverse quantization | 36 |
| (b) Inverse DCT | 240 |
| (c) Prediction+Differences | 124 |
| **TOTAL** | **4908** |
| **DECOMPRESSION** | |
| Entropy Coding - Decoder | 68 |
| Inverse Quantization | 36 |
| Inverse DCT | 240 |
| Loop Filter | 0 |
| Prediction | 180 |
| YCrCb to RGB | 108 |
| **TOTAL** | **632** |

## 4 Television Computing & the Consumer Appliance Vision

One can anticipate a broad range of consumer devices in future interactive video systems serving a broad range of consumer needs at a cost that is mandatory for mass marketing. Such appliances can be classified into four classes, namely Information-centric (PC, MAC), Entertainment-centric (TV, Settop), Communication-centric (Phone, Fax) and In-Home Information-centric (Security, Power). Issues such as ease-of-use, plug-and-play and interoperability are paramount. The notion of "Television Computing" in the home as opposed to "Desktop Computing" in commercial settings is oriented towards mass-audience, effortless interaction and immediate response where the communications model is broadcast. This environment is more screen oriented than window oriented and the visual and audio expectations are high.

### 4.1 The Settop Interface

The set-top device has two interfaces[6]. On one side, the set-top connects to a digital/analog communication channel. This channel connects the set-top to an information infrastructure such as Level-1 gateways or "Head-Ends", Level-2 gateways and to servers. On the other side, the set-top interfaces to a user through display devices such as a television and input devices such as a remote controller. The set-top obtains a multi-modal data stream from the network consisting of, for example, digital video, digital audio, images, user-interface components and graphics. The set-top can also generate a multi-modal data stream to put on the network comprising of the same components, but initially probably mostly data.

We expect that the mode of services in full service networks will work as an extension of the model of current cable networks. In the current cable systems, there are a number of channels that are available to the user as basic service. In addition, there are certain channels from which the user can order specific programs (e.g. Pay-Per-View). In addition, there will be interactive channels that will provide services such as video-on-demand, games-on-demand, news-on-demand and shopping-on-demand. These channels require some interaction on the users' part to benefit from the provided services.

### 4.2 The Settop as a Computer

Figure 3 shows a comprehensive architecture of a set-top. In essence, a set-top box architecture looks very similar to a multi-media capable computer. Like a computer, a set-top box has a CPU, memory, graphics and peripherals. In addition, there are powerful digital video and audio capabilities. It is important to note that these audio-visual capabilities are much more powerful than most of today's computers. The most powerful of computers can barely play SIF resolution (352x240) at full motion rates (30 frames/sec). This resolution is comparable to a VHS quality tape recording. Consumers on the other hand expect a visual quality comparable to broadcast quality which requires a resolution of 720 by 486. In addition, the audio quality expectations are comparable to CD quality which requires a resolution of 16 bits per channel at data sampling of 44.1 Khz or better and surround sound. FM synthesis and "SoundBlaster" capabilities may also be expected.

Another important difference between a computer and a set-top box is related to security and authentication. As opposed to a computer, the primary function of a set-top box is to enable subscription services where a user pays for services that he/she uses and content access. It is very important for both the service providers to be compensated for services that they provide and for consumers to be fairly charged only for what they use. Traditionally, providers of cable television service companies have spent considerable effort to develop their security systems to prevent unauthorized use of their services[7]. The cable providers (also referred to as Multiple Service Operators or MSOs) and other service providers will require that their investments be protected and fraud mitigated.

There are some limitations of Television technology also. One has to deal with an interlaced NTSC resolution as opposed to higher resolution computer displays such as Super VGA. The inherent bandwidth limitations of NTSC impose some quality limitations especially in the text area. NTSC was optimized for continuous moving pictures and does not perform so well to display sharp edged stationary objects like text. Similar limitations exist in the realm of color fidelity and bandwidth.

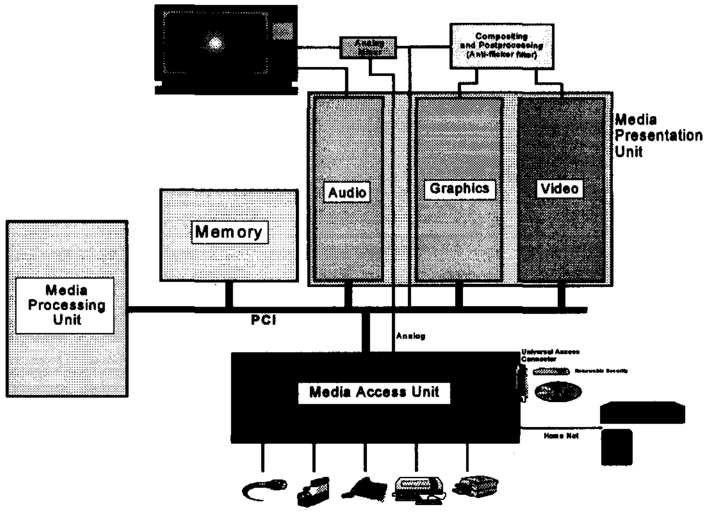

Figure 3 - Simplified Settop Architecture

### 4.3 User Interface Models

Considering the installed base of televisions in homes the most popular interface to the set-top seems to be a television. A set-top produces analog video signals comprising audio and visual components as directed by the incoming data stream and user interaction. This analog video signal is a composite of digital video streams from a server and graphics possibly generated locally.

The basic hardware for a user to interface to the set-top is a simple infra-red remote control, with a small number of buttons. Another possibility for an interface model to a full-service network is a personal computer through an augmented modem to connect to a cable. This cable-modem may be able to share some of the resources in the computer such as CPU, memory and peripherals. This mode provides more latitude and freedom for interface design because of a multitude of input devices, such as a keyboard, and the higher spatial and color resolution of display. On the other hand, the TV interface is, at least currently, more prevalent and convenient for a user.

## 5 Hardware Architecture

The set-top interfaces to the network through a tuner, demodulation and descrambling module. There is also available a reverse channel, most likely, of a lower bandwidth than the down stream channel. A media processor module provides video and audio processing capabilities. A graphics module provides local graphics capabilities to produce user-interface and other capabilities. An analog module converts digital video and audio signals to be fed to a display.

### 5.1 Demodulation and Transport

The set-top interfaces to the network through a transmission medium which delivers the content for the set top. The transmission interface would be a co-axial cable or fiber optic channel. Some companies are also investigating wire-less technology to deliver data to the home. It seems that, at least in the near future, fiber to every home would be too expensive a proposition. The most like scenario for the near future would be fiber to the curb. From there, a coax connection would be run to each home.

The transport protocols are also an interesting issue. In the long term, an ATM/TAXI/SONET based protocol may be used to transport data. Currently, however, interfaces for these protocols are prohibitively expensive for a consumer application or deployment. In the short term and until a standard emerges, proprietary data transport protocols will be used. Issues such as QAM Vs. VSB modulation, MPEG2 transport Vs. ATM transport, cascading of error correction techniques and security are all presently under the control of the network provider.

### 5.2 Video and Graphics Processing

MPEG2 as a digital video transport standard has been gaining popularity in industry and is the most likely candidate for a set-top box. Some cable companies have developed proprietary digital video transport protocols and they might coexist in the near term. Discussions are underway also to standardize on digital audio standards such as Dolby's AC3, Musicam or MPEG. The decoding of an MPEG2 video stream requires about 800-1000 MOPS . To support this computational requirement, dedicated processing hardware would be required in the near term. A number of companies including C-Cube, Philips, AT&T, LSI Logic, Hyundai, Samsung and SGS Thomson have announced chip sets for MPEG2 decoding.

Besides MPEG1 and MPEG2 decompression, an advanced set-top box will also support:

- **Compositing 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 subimages 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. Another application is picture-in-picture display. 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? 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] gives a low noise result with a computational advantage of a factor of five (5276 ops Vs 880 ops for each output 8x8 block). Other operations such as editing in the compressed domain or filtering are also of interest. The compositing function might be enhanced in the future to include mixing of graphics and video streams as well for decoding by the settop.

- **Object tracking & Graphics/Video integration** - 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 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 4 below 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. Such operations will enhance the viewing of digital video in the future and can be done in conjunction with the settop appliance.

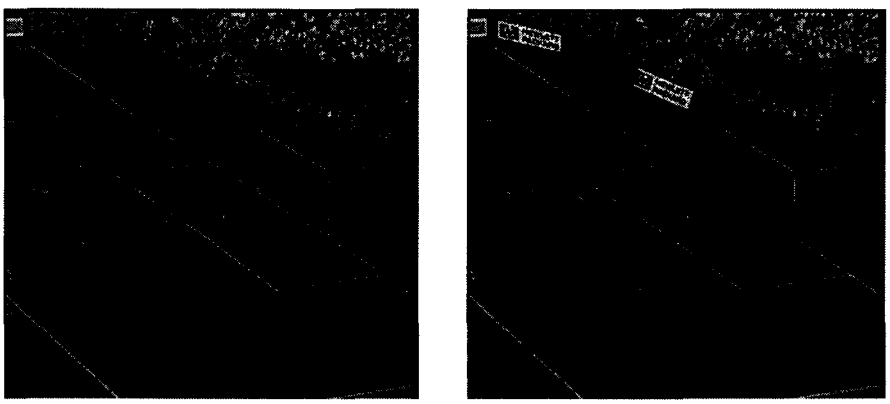

Figure 4 - MPEG2 Resolution Video Frame with Texture Mapping

- **Resolution Conversion.** In order to support an enhanced display, some resolution conversion might have to be performed on the settop. Furthermore, if the user desires to print the incoming video, often, deinterlacing and scaling of the video is needed in order to get a high quality printout on a 300-600 dpi print device. The functions used in resolution conversion are essentially the same functions used in object tracking; however, the granularity of the functions in the former case is at a higher resolution

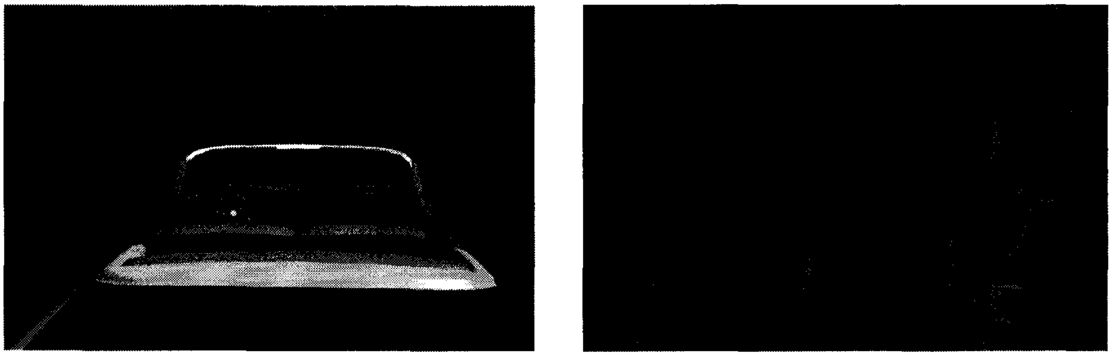

Figure 5 - a) 3D Graphics & Image Composite b) N-Dimensional Video

**Graphics.** Graphics would be required to present user interface elements for navigation and presentation. In addition, an important role of graphics would be to support interactive graphical applications such as interactive games. One of the applications envisaged for a set-top box is games-on-demand. A video game would be downloaded into the set-top box to be played. The graphics hardware in a set-top box would have to support 2D and 3D graphical elements such as lines, fills, patterns, and textured-mapped shaded polygons. High performance graphics, both two-dimensional and three-dimensional will be provided. Two-dimensional graphics is necessary for basic user-interface elements. Two-dimensional graphics support will exist in the form of hundreds of thousands anti-aliased vectors and polygons per second, hundreds of sprites, and anti-aliased text. This will enable animated and colorful gripping interfaces. Three-dimensional graphics will support advanced navigational systems, games and new applications such as home shopping. Performance will be of the order of a quarter of a million of fully shaded, lighted and textured polygons. A set-top box will be able to composite graphics and digital video to create engaging and interactive applications as illustrated in Figure 5a above where a 3D car model is composited with a 2D image background..

Games may be categorized into different categories according to their resource requirements. Some of the resource categories are: latency, bandwidth, 3D graphics, 2D graphics, storage and computation. Storage will be limited in the beginning but the set-top may be able to rely on the server for some of its transient storage needs and multi-user communications. An example might be as demonstrated in **Figure** 5b where many images of a 3D object with scale and rotational changes would be stored on a server for a home shopping application. The user at the settop would experience an interactive exploration of such an object with merely changing the sequencing of the animation frames at the server, based on user input, without using advanced graphics in the settop.

### 5.3 Central Processor Unit and Media Processors

A central processor unit and associated memory will provide basic control functions in the set-top box. The integration of video and graphics processing with the CPU would form 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 PA7100LC processor design may be enhanced with specialized instructions or co-processors. 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. 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.

### 5.4 Complexity Analysis

The key to the mass deployment of the settop box will be the integration of most key functions into VLSI. As indicated in Figure 3, the settop architecture can be logically divided into a media processing (video, graphics & audio) section, a communications (adaptive equalizer, QAM demodulation, Reed Solomon/Viterbi error correction) section and a control (interface and graphics). Both graphics and video may be supported by the processor or specialized coprocessors. Recent work reported by Chatterjee[9] is (shown in the top) indicates an approximate gate budget for the various components of the settop as shown in Table 4.

Table 4 : Digital Cable Settop Complexity (Chatterjee)

| Function | Gate Count | RAM (Bits) | ROM (Bits) | Transistors |
|---|---|---|---|---|
| 64 QAM | 70K | | | 280K |
| Viterbi | 35K | 27K | | 302K |
| Reed-Solomon | 30K | 18K | | 228K |
| Transport | 50K | | | 200K |
| OSD | 4K | 4K | 32K | 70K |
| MPEGII Video | 117K | 50K | 32K | 800K |
| MPEGII Audio | 34K | 44K | 52K | 450K |
| | | | | |
| Processor | 40K | 4K | 200K | 500K |
| | | | | |
| **TOTAL** | **380K** | **147K** | **316K** | **2.83M** |

## 6 Software Architecture

For a programmer, the set-top provides an API at multiple levels.. At present, the operating system decision is usually closely coupled with the processor choice. In the most likely scenario, which provides a high level of openness and expandability, application code will be downloaded to the set-top. This scheme provides an open interface for application developers. The division of an application between the set-top and the server is an interesting issue. On the other hand, there will be powerful servers connected to the set-top. A good application will utilize the resources of both the set-top and the server judiciously.

In one scenario, an application may be running on the set-top interacting with the server. Video-on-demand is an example where the media player is running on the set-top interacting with the server to get video. Figure 6 shows an overall system for such a Video-on-Demand system. Most applications will have a set-top component and a server component. The application will define how these components communicate. The set-top will provide a basic messaging API to enable this communication. In terms of data formats, video and audio would probably be encoded as an MPEG2 stream. Proprietary schemes may have to be supported also, at least in the short term. Encoding for other streams such as overlays, graphics, access and control would be devised. It is possible and work is underway to devise schemes to package this data as part of an MPEG2 stream.

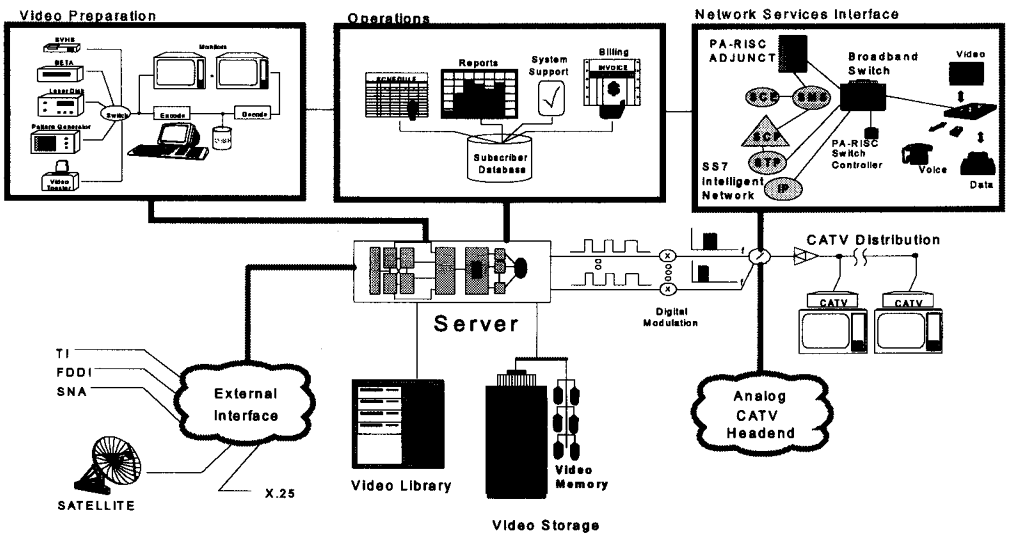


Figure 6 - Video-on-Demand System

## 7 Inter-operability Issues

For each of the elements of the full-service network (e.g. services, servers, networks, and set-tops), there will be a number of vendors. Each of these components should be able to support a number of different device types to which they connect to allow consumers to move equipment, for example, and to enjoy the cost advantages of volume. For example, each set-top has to be able to support multiple video information providers (VIPs). Similarly, each VIP should be able to provide its video information on set-tops from multiple vendors. It is imperative for standards to emerge if equipment and services are to be affordable to both providers and users.

## 7.1 Application Programmer's Interface

To accelerate the early creation of content to fill up the hundreds of megabits of channel capacity, the content providers should be provided with an application programming interface (API) to which they can develop applications. An API also serves to abstract hardware. Of course, vendors of servers and set-tops will provide APIs for their platforms and a de-facto standard may emerge. It would be more expedient if an industry forum developed and accepted a mutually agreeable standard. An object-oriented approach that seems to be getting some attention is that of the Object Management Group (OMG) with their interface definition language (IDL) and Object Request Broker (ORB). This API has to support a number of very different applications such as video and audio play, video games, home shopping, and digital editing.

# 8 Conclusion

The efficient support of interactive video requires an intimate knowledge of video algorithms and available computing architectures. The opportunity exists to leverage the high volume and demanding requirements of settops to "high performance" workstation multimedia computing. The possibility exists for interactive video client-server applications to span from commercial systems to consumer systems. This connection between "work and play" will ultimately engender the biggest impact on multimedia deployment as it provides value to communications, education, commerce, health care and entertainment.

**Acknowledgment**

The MPEG1 video player work is based on work across various entities within HP. The authors wish to acknowledge the efforts of Ruby Lee, Peter Kaczowka, Pat Mcelhatton, John Beck, Behzad Razban and the other members of the Multimedia Architecture Team and the MPower team. Other contributors to this research include Balas Natarajan and Konstantinos Konstantinides at HP Labs. Several Divisions of HP are contributing to the vision of the settop presented here and include the Home Products Division (Webb McKinney, John Alburger) and the Video Communications Division (Jim Olsen, David Andersen, Andy Gee).

**References**

[1]P. Knebel, B. Arnold, M. Bass et. al, "HP's PA-7100LC: A Low-Cost superscalar PA-RISC processor", Proceedings of IEEE Compcon, pp. 441-447, Feb. 22-26, 1993, San Francisco, California.

[2]ISO/IEC JTC1 11172, "Information Technology - Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mb/sec - Part 2", Coding of Moving Picture Infomation. International Organization for Standarization, 1991.

[3] R. Lee, "Multimedia Enhancements for PA-RISC Processors", Hot Chips Symposium VI, July 94, Stanford University, California.

[4]A. Barkans, "Color Recovery: Millions of Colors from an 8-bit Graphics Device", Proceedings of the IEEE Compcon, March 5-9, Feb. 1995, San Francisco, CA.

[5]K. Konstantinides, V. Bhaskaran, *Recent Developements in the Design of Image and Video Processing Ics, Chapter 2 - VLSI Signal Processing Technololgy*, Kluwer Academic Press, 1994.

[6]B. Furht, D. Kalra and F. Kitson, " Design Issues for Interactive Television Systems", Computer, Vol. 28, No. 5, May 1995, pp25-39.

[7]*Video Scrambling and Descrambling for Satellite & Cable TV*, Rudolf F. Graf and William Sheets, Howard W. Sam & Co., Indianapolis, 1986.

[8]B. Natarajan and V. Bhaskaran, "A Fast Approximate Algorithm for Scaling Down Digital Images in the DCT Domain", patent application pending.

[9]P. Chatterjee, Gigachips: Deliver Affordable Digital Multimedia for Work and Play Via Broadband Network and Set-top Box, IEEE ISSCC95, pp26-30.