Survey of Personal Interactive Multimedia Technologies

Chingshun (Jesse) Cheng
IBM Corporation
Advanced Workstation Division
Austin, Texas 78759

C. Jinshong Hwang
Department of Computer Science
Southwest Texas State University
San Marcos, Texas 78666

Abstract

This paper surveys current personal interactive multimedia technologies. The paper first describes background technologies such as image and audio technologies. Secondly, a description of hardware and software architectures that involve multimedia is given. Finally, current major implementations of personal interactive multimedia technologies such as Intel's DVI, IBM^{R}'s AVC, Apple's HyperCard, and Commodore's Amiga are described. The future outlook for multimedia technologies is briefly discussed.

1 Introduction

"Multimedia" implies multiple media. Unfortunately, there is no standard definition for the word "multimedia" in industry at this time. For the sake of clarity, this article uses the word "multimedia" to convey the coordinated use of multiple media – typically from a group which includes text, graphics, animation, sound, full motion video or program. In addition, the word "multimedia" in this article has the connotation of "interactive". That additional feature comes directly from the computer that controls the multimedia. Obviously, a regular movie is excluded as multimedia here due to the lack of interactiveness.

The multimedia system can be used to develop various multimedia applications. Examples of these include training [1], teaching [2],game playing[3], showing properties in the real estate business, providing multimedia database services in the medical field [4], and giving multimedia presentations to management.

The word "hypermedia' used in industry also deserves some attention. Hypermedia is a form of multimedia. It utilizes information under the control of a computer so that the user of the information can navigate through it effectively. The information may be in the form of texts, diagrams, moving diagrams (animation), images, moving images (television), speech, sound, or computer programs [5]. The user can easily navigate thru different display information using the buttons on the display. He can also reverse the action of the button to return to the previous screen. In this article, the word "personal" is emphasized. It is chosen to better focus on the technology that involves only personal multimedia presentations. Technologies that involve CD-ROM etc. will not be considered in this article due to the fact that the regular user has no way of storing personal information on those storage media.

Typical multimedia systems involve at least three separate technologies. Namely, realistic image technology, digital audio technology, and computer technology. This article will cover the basics of realistic image and digital audio technologies before introducing typical PC and personal multimedia system architecture.

2 Realistic Image Technology

Currently, most of the cameras use a CCD (charged coupled device) for image capturing and signal digitizing. Charged-coupling involves the collective transfer of all the mobile electric charges ("packet") stored within a semiconductor storage element by the external manipulation of voltage [6]. The device generates the analog signal by rippling the charge-forms in the cells constantly. The analog signal outputs from the CCD sensors is converted into RGB (Red,Green,Blue), Y/C (Luminance/chrominance, also called S-video or S-VHS), or NTSC (National Television System Committee) so that the signal can be transmitted further. In the case of both the Y/C and NTSC formats, the three colors are funneled into two wires (Y/C) or single wire (NTSC), and, therefore, the color resolution is poorer. The Y/C or NTSC signal format is not recommended for the highest quality work.

The number of bits assigned to describe each picture element depends on the complexity of the image. Signal digitization is done through sampling and quantization (i.e. analog-to-digital conversion).

The digitized image data can then be transmitted into the system memory at the PC or the random access memory at the display adapter card. The display is created by continuously reading the bit stream from that memory, converting it to a video signal, and displaying it on a raster-scan (or simply raster) display.

Most people measure noise in terms of signal- to-noise ratio, which is defined for images as the ratio between the peak-to-peak black-to-white signal and the rms (root-mean-square) value of any superimposed noise. A good monochrome system should have a S/N (signal-to-noise) ratio of 46 dB.

It can be shown that at least 5 bits are required to get a reasonable picture. In reality, experience shows that 5 bits are not enough to get a clear picture due to the digital artifact – contouring.

For a color system, it is possible to create quite realistic images using 16 bits per pixel (bpp) by compromising the number of bits allocated for each color.

Techniques to reproduce good images with fewer bits than the above number are called video compression techniques. Compression/decompression is usually accomplished by utilizing special hardware to free-up CPU cycles. Through compression/ decompression, one can drastically reduce the amount of data needed to be saved in storage for image reproduction. However, the compressed data has to be decompressed before any image manipulation can be done. It requires re-compression again after the manipulation.

An alternative use of a bitmap to represent an image is to use pixel values as an index into a color look-up table (LUT) which holds the actual color value. The LUT is loaded with the application software. There is no overhead involved in LUT systems in terms of manipulating the compressed data. Therefore, less bpp can be used to achieve high quality image representations and high efficiency image manipulations.

3 Digital Audio Technology

Audio production is typically done with microphones and tape recorders. Audio recorders for production usually have one or two tracks on magnetic tape. In broadcasting or professional level productions, the audio is recorded separately and put through separate post-productions. Various audio recording and reproduction techniques can be found in Bruce Bartlett's book about professional recording [8]. Only at the end of the process will the audio and video be combined on the same tape.

The same analog-to-digital conversion applies to digitizing analog audio signals except that the S/N ratio in the latter case has to be at least 96dB for noise not to be heard in normal listening.

Quantizing of the 16-bit digital audio is normally done with linear PCM (Pulse Code Modulation), where each step of quatization is of equal size. This is the format used with the audio compact disc. However, with fewer than 16 bits per sample, other strategies such as Nonlinear PCM, differential PCM, or adaptive differential PCM are worthwhile. More detailed information on the above can be found in books such as Arch C. Luther's Digital Video in the PC Environment [7].

Many personal computer operating systems (such as MS-DOS) do not provide real-time operations. However, one can use the interrupt capability to synchronize various audio and video activities on system bus. The audio data is transferred in blocks while the rest of the application (video and computer function) can be designed so that an access to the hard disk is achieved thru processor interrupts.

4 Audio/Visual/PC Personal System Architecture

4.1 Hardware Architecture

Figure 1 shows a simplified block diagram for a PC/AT or compatible computer systems. The system bus in this system is used as a common channel between CPU and various I/O devices or memory. The system bus is capable of a 16 bit data transfer and 24 bit addressing to/from system memory. The maximum processor speed is typically 2.67 million cps. In addition, the processor contains the system clock and also handles inter-

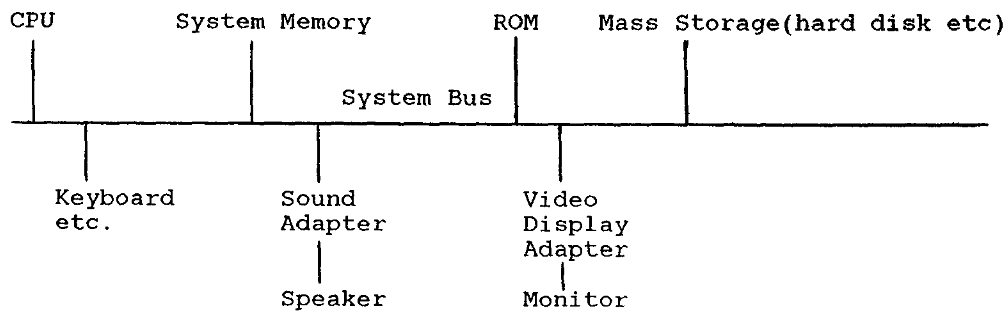

Figure 1: Typical PC/AT Computer System in Block Diagram

rupts.

Realistic video images for display in a desktop situation usually require at least a pixel resolution of 512x480 or higher and a bpp value of at least 16 bits for non-LUT [7]. (AVC uses 8 bits and achieves very good quality due to LUT implementation.) This means that a single full-screen image takes 491,520 bytes. In the presence of the MS-DOS operating system, the typical hard disk data-rate is less than 150,000 bytes/sec. Thus, 3.3 secs. will be required to retrieve the full-screen image from hard disk, even if we write from hard disk directly into the screen memory of a realistic video display adapter. Obviously, it will help if we deal with less data for a frame video. That objective can be achieved by: (1) compressed images and include a very fast decompression capability in our system; (2) reduce the size of a motion window until we can handle the pixel rate; (3) combining both (1) and (2) above. A realistic video system design should allow the flexibility of combining methods to achieve motion video in any size window and with any kind of compression. The same thing applies to audio, so we can tailor the use of the resource data-rate dynamically within an application.

Due to the limitation of CPU processing speed, data operations can best be carried out by powerful and programmable co-processors. In essence, beyond screen display memory management and screen refresh, the hardware needs the implementation of the best algorithms for various audio and video data manipulations.

4.2 Software Architecture

Typically, the application software interacts with the high-level modules such as the graphics support system and the audio/video support system. The graphics support system provides image handling graphics, drawing primitives, and video manipulation functions. The audio/video support system is used for playback of the audio/video file from hard disk, CD-ROM, or RAM. These high-level modules are typically interfaced with audio/video adapter cards thru driver interface modules and various device drivers. These drivers are not accessible to application programs except through the driver interface modules which sit above the drivers. The drivers are loaded when the computer is booted up and remains in system RAM. The rest of the runtime software is contained in the system libraries, which becomes part of an application program thru linking [7].

The vast majority of available multi-media software use the "hypermedia" concept. According to R. M. Akscyn et al. [9], hypermedia is a generation of the hypertext concept now over 40 years old. Most hypermedia and hypertext systems are characterized by Akscyn et al. [9] as follows:

- Information is "chunked" into small units. Units may contain textual information. In hypermedia systems, units may also contain other forms of information such as vector graphics, bitmapped images, sound and animation. These units are named notecards, frames, nodes, etc. in various hypermedia systems.
- Units of information are displayed one per window (Systems vary in the number, size and arrangement of windows they permit).
- Units of information are inter-connected by links.
- By creating, editing, and linking units, users build information structures for various purposes (e.g. authoring documents, developing on-line help systems, etc.)
- In shared hypermedia systems, multiple users may simultaneously access the hypermedia data base. Shared systems may be implemented as distributed systems, in which portions of the database are distributed across multiple workstations and file servers in a network.

The link starts with "button" and ends with "target". (Figure 2)

A button is usually represented by either a phrase or a marked area in the text or diagram. It is commonly highlighted through the use of reverse video, color, or pop-up box. The user may activate the button through various input/output devices.

The target mentioned above could be a line or segment of the same text, another document, a picture or diagram, a moving video or animation sequence, or it could even be a program.

When the button is activated, it results in the actions prescribed at the target. An example of this is AVC which has been developed by IBM.

5 Still Images and Special Effects

Compared to full motion pictures, still images are different and equally powerful in conveying information. A decisive moment usually requires the still image so that the audience can concentrate only on key information and study it. Also, the layout of multiple still pictures can often be used to organize pictures spatially and allows the user to participate in the story in a more efficient way. An analogy to this is the National Geographic magazine vs. the National Geographic explorer. They coexist well due to their individual values.

The goals for still picture presentations are to achieve better picture quality, to use less storage, and to be able to achieve image retrieval in a short time [7]. In principle, the still image hardware organization can be shown in Figure 3 for a typical PC.

The video/audio capture equipment can be a video-camera, a tape recorder, or a scanner. The analog signal will be converted into digital signals before interacting with the system bus. On the other hand, special video cameras such as the telecine camera is used to obtain high-quality video reproduction from motion picture film or slides.

Special effects usually include fades, dissolves, wipes, or overlays etc. They are accomplished by specially designed software and hardware that handle various types of image manipulation functions.

Examples of how some of the transition effects can be achieved are described in Arch C. Luther's "Digital Video in the PC Environment" [7]. The book also briefly shows how operations which change the value of pixels (i.e. image processing operation) can be done.

6 Current Major Implementations of Video/Audio/PC Personal Systems

6.1 INTEL'S DVI Technology

Intel's DVI (Digital Video Interactive) technology combines the interactive graphics capabilities of personal computers with the realism of high quality motion video and multi-track audio in an all-digital integrated system.

DVI is the work of a pair of chips: the i750 chip set. One is the 82750PA Pixel Processor, a 12.5 MIPS chip (an AT runds at about 2 MIPS) with its own parallel processing architecture and on- chip RAM for quick reprogramming. The other chip is the 82750DA Output Display Processor, which can serve up to several different resolutions, from 256 x 200 pixels to 1024 x 512 pixels in 8, 16, or 24- bit color. Those chips are able to decompress full-screen motion video in real time as well as to accomplish interpolation of the U and V color components for each pixel in PAL (Phase Alternation Line) format.

Another key successful factor of DVI technology is the multiple step compression technique. According to a paper written by G. David Ripley [12]:

> ....each 3/4 M byte image must be reduced to around 4500 bytes, a compression factor of over 160. DVI represents the pixels in the PAL YUV (Y component is the liminance component in PAL) in format instead of the more typical RGB graphics format. This scheme permits a reduction in image size while retaining surprisingly good image quality. By only updating U and V every four pixels horizontally and vertically, and interpolating their values at run time, the average pixel depth is reduced from 24 bits to 9 bits, or nearly a factor of three [13]. A further compression factor of four is achieved by simply decompressing images to a resolution of 256 by 240, a factor of 2 in each direction less than that of the original digital video. Finally, a sophisticated compression algorithm analyzed image correlation, both intra-and interframe, resulting in an image data reduction of a factor of 10 to 15 or so, depending on the amount of action within the scene.

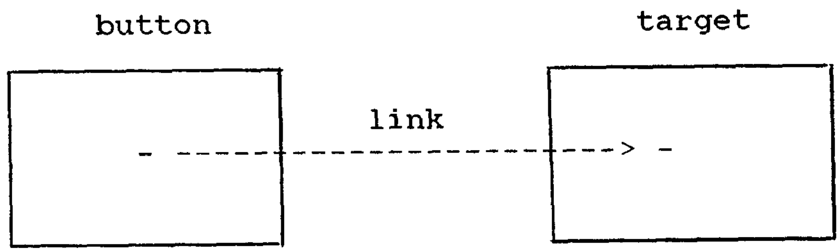

Figure 2: The Relationship Among Button, Link, and Target

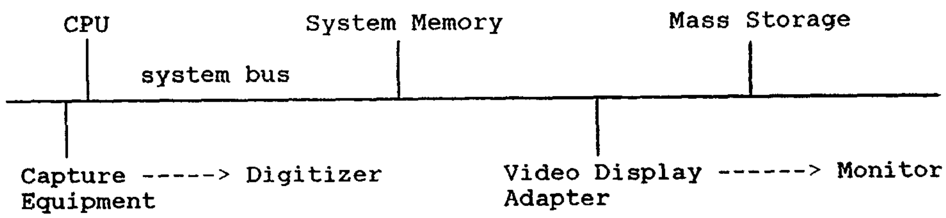

Figure 3: Typical Still-Image Hardware Organization

6.2 IBM's Audio Visual Connection (TM) Technology

AVC is basically implemented with the video capture card, audio adapter card, motion picture adapter card, as well as PC. Currently, there is no hardware available that allows the user to compress/decompress the image data. It represents state-of-the-art low cost personal multimedia presentation system from IBM.

Alfred Poor [15] made the following statements about AVC:

> "AVC give you the tools to combine brilliant color, animated images, life-like video-pictures, an synchronized voice andmmusic sound tracks into a unified presentation....It is the first (product) to offer a turnkey solution for producing and replaying multimedia presentations on the IBM PC."

AVC is basically a manually driven program that allows the user to do the following:

- edit graphics and text
- create and edit audio objects
- record sounds and add them to a story
- capture images and edit them for use in audio-visual shows
- full-motion video pictures to a story
- convert images
- reference applications and objects

It utilizes the on-line hypertext/graphics/ audio-help system. For advanced users or expperienced programmers, AVC authoring language (AVA) provides flexibility in terms of adding more advanced features to a story such as file input/output, logic and branching, trigger fields, and variable text input and output. Programs that involve artificial intelligence and AVA are possible. They can be used to cover a wide range of applications. AVA also allows the user to manage host computer connections, integrate C programs and call on the special capabilities of OS/2TM. Due to the relatively inexpensive image data manipulation hardware, the cost of AVC is quite a bit less than systems using DVI technology.

6.3 APPLE Macintosh's Dekstop Media Technology

Apple's introduction of the HyperCard has played a very important role in focusing attention on multimedia.

Apple's aggressive promotion of it is changing hypertext from an esoteric concept known to only a few hundred people, to a household staple of computing used by millions [16].

As indicated by Gregg Williams [17], think of HyperCard as a stack of 3 x 5 inch index cards. According to Apple's terminology, a group of related cards cards and the file that contains them are collectively called a stack. According to Salomon et al. [18], the goal is to create a data base which demonstrates delivery of multimedia reference content on an economical PC platform. According to the same article [18], the data base exists as several "stacks" of HyperCards. Each card is equivalent to a screen, and only one card can be viewed at a time. The mouse is used to click the buttons to take certain pre-determined programmed action while the cursor is within one of these buttons. As is in Hypertext, Links are used among cards for forward and backward references. HyperCard achieves forward and backward references. HyperCard achieves animation the same way as it achieves story-telling. By clicking the button representing the desired image sequence, the user can view the image within the sequence either individually or as an animated sequence.

Atkinson designed Hypercard so that it appears simpler to novice users and powerful to sophisticated users (e.g. browsing, typing, painting, authoring, and scripting are for everybody and meta scripting is for the programmers). According to Gregg Williams [17]:

> "You write the appropriate code in C, Pascal, or assembly language, compile it without a header as a code resource of type XCMD (for message handlers) or SFCN (for functions), and paste it into that resource's name in HyperTalk scripts, Hyper- Card executes the XCMD or XFCN resource with the same name."

Apple has yet to develop a data compression/decompression technology for video data [10]. However, Apple does have an algorithm that allows a 30x-size compression of graphic images [18].

6.4 COMMODORE AMIGA's Multimedia Technology

When Commodore talks about multimedia, it emphasizes the desktop video. However, the relatively inexpensive AMIGA hardware and its powerful authoring system (AmigaVision) integrates graphics, animation and sound with its multi-tasking operating system.

Amiga's hardware implementation involves three custom chips that Commodore developed: Agnes, Paula, and Denis. According to Bob Ryan [19], Denis can support 640 x 400 pixels without any additional hardware. Paula provides the Amiga's four-voice sound. Agnes allows chip RAM access up to 2 megabytes. (In Amiga's terminology, chip RAM is used primarily to hold graphics and sound data for both CPU and custom chips. Fast RAM is used in the same fashion as normal system memory. Only CPU is allowed to access the executable code and non- display data in fast RAM). Although Amiga is equipped with superior graphics performance, Amiga systems such as Amiga 3000 can only support 32 colors in low resolution and 16 colors in high resolution, out of a possible 4096. No known techniques are currently available to improve the situation without getting performance hit. Amiga needs to provide native 8- and 24-bit color modes in order to remain competitive in the market.

Amiga's OS 2.0 operating system offers good user interface. It provides Amiga with good multi- tasking ability.

7 Conclusions

Many companies (such as IBM, Intel, Microsoft and Lotus) have projected that between 1992 and 1995, the "plain vanilla" PC sold will contain built-in multimedia capability [11]. As indicated in this article, technology such as DVI is just a step away from full-motion, full-screen video. On the other hand, technology based on the hypermedia approach will definitely gain additional strength in terms of artificial intelligence and communication developments. Recent publications indicate that the multimedia relational data base has been successfully designed and implemented [14,20].

Many data base researchers are looking forward to developing the multimedia object-oriented data base that involves artificial intelligence and communication.

Many companies view an early involvement in this technology's evolution as vital to their companies' future success in the personal computer market place [11]. The dream of good and affordable personal interactive multimedia is coming true.

8 Acknowledgements

The authors would like to thank Dr. Albert Edgar at IBM Austin for technical advice and fruitful discussions. The authors would also like to thank Mr. Wayne Galella at IBM Austin Multimedia Lab for providing the environment to conduct this research project.

REFERENCES:

1. W. D. Kahri, "Comprehensive Program Trains SP and C Technicians", Electr. World (USA), Vol.3, No.2, Feb. 1988.
2. L. Friedlander, "Moving Images into Classroom Multimedia in High Education", Laserdisk Prof. (USA), Vol.2, No.4, Jul. 1989, pp. 33-38.
3. A. Druin, "Noobie: The Animal Design Play- station", Sigchi Bull. (USA), Vol.20, No.1, Jul. 1988, pp. 45-53.
4. Abul Huda et al., "Multimedia Data Bases in Medicine", European Federation for Medical Informatics. Proc. of the Seventh International Congress, vol.2, 1987, pp. 915-921.
5. James Martin, "Hyperdocuments and How to Create Them", Prentice-Hall, 1990.
6. Fairchild, "Charged Coupled Device (CCD) Catalogue", 1982-1983.
7. Arch C. Luther, "Digital Video in the PC Environment", Intertext Publications, McGraw- Hill, 1989.
8. Bruce Bartlett, "Introduction to Professional Recording Techniques", Knowledge Industry Publications, Inc., 1987.
9. R. M. Akscyn, et al., "KMS: A Distributed Hypermedia System for Managing Knowledge in Organizations", Communications of the ACM, Vol.31, No.7, Jul. 1988, p. 820.
10. Phillip Robinson, "The Four Multimedia Gospels" BYTE, Feb. 1990, p. 203.
11. Dwright B. Davis, "Intel and IBM Share Their Multimedia Vision", Electronic Business, Nov. 13, 1989, pp. 26.
12. G. David Ripley, "DVI - A Digital Multimedia Technology", Communication of the ACM, Vol.32, No.7, July 1989, p. 811.
13. M. Tinker, "DVI Parallel Image Compression", Communications of the ACM, Vol. 32, No. 7, Jul. 1989, pp. 844-851.
14. Cathy Anne Thomas, "A Program Interface Prototype for a Multimedia Database Incorporating Images", MSCS Thesis, Naval Postgraduate School Dec. 1988.
15. Alfred Poor, "Turnkey Multimedia from IBM," PC Magazine, May 15, 1990, p. 158.
16. John B. Smith et al., "Hypertext", Communications of the ACM, Vol.31, No.7, Jul. 1988, p. 816.
17. Gregg Williams, "Hypercard"' BYTE, Dec. 1987, p. 109.
18. Gitta Salomon, Tim Oren, and Kristee Kreitman, "Using Guides to Explore Multimedia Databases", IEEE, 1989.
19. Bob Ryan, "Commodore Sets Course for Multimedia", BYTE, May 1990, p. 122.
20. Tim Shelter, "Birth of the BLOB", BYTE, Feb. 1990, p. 221.