Professional Video Under 32-bit Windows Operating Systems

By Alain Legault and Janet Matey

The personal computer (PC) and 32-bit Windows operating systems are quickly becoming the platform of choice for professional video, film, and multimedia producers. Windows users have long benefited from the time and money savings offered by "open platform" interoperability. Professional media producers need the same opportunity to be able to buy the best interoperable tools, from an assortment of vendors, at a variety of price/performance points to meet their individual requirements. Many leading video hardware, software, and system vendors are committed to delivering video and audio editing, multimedia authoring, plug-in effects, and animation and graphics creation applications that all work together on PCs with varying performance capabilities. The goal is to let users choose all the best tools for their jobs at competitive prices.

This paper describes a software architecture for professional video applications based on the enhanced AVI file format, the Windows NT file system (NTFS), and ActiveMovie. It also proposes a set of enhancements to ActiveMovie that will ensure that manufacturers are able to deliver interoperable, broadcast-quality production tools on 32-bit Windows PCs.

PCs began to look like useful tools for the video world at the end of the 1980s with the introduction of the first multimedia products. At that time it became possible for vendors to satisfy certain segments of the emerging multimedia market with the technology available: 386 processors, Windows 3.0, and ISA/EISA expansion slots. Professional video and film producers, however, could make only limited use of these early tools, because the performance of available hardware and software architectures did not allow broadcast-quality, field-accurate editing.

System Resource Requirements for Professional Video Applications

Over the last few years, technology has evolved to the point where the PC has become a significant force in the market for professional video applications, particularly in digital nonlinear video editing systems but also in digital disk recorders, broadcast media servers, live video switchers, M-JPEG-to-MPEG transcoders, commercial insertion systems, CATV barker channels, CG/paint/animation workstations, and visual effects/compositing systems.

At NAB '96, over 80 Windows-based digital nonlinear video editing products were shown and numerous examples of the other types of systems were also prominent. The many advantages of the PC in professional video applications were clearly demonstrated — cost effectiveness, high performance, networkability, and the availability of a wide variety of application hardware and software. In the next few years, innovations and improvements in both hardware and software technologies will continue to improve the performance of the PC in the professional video marketplace.

Hardware Architecture

PC solutions for professional video users are typically based around a single video adapter or a set of adapters that plug into PC expansion slots. These expansion cards, equipped with specialized processors (application specific integrated circuits [ASICs] and digital signal processors [DSPs]), perform high-quality algorithmic processing at real-time video speed; functions that cannot be executed by the host processor alone. Functions of the dedicated cards include the following:

- Video acquisition and playback.
- Audio acquisition and playback.
- Video algorithms (i.e., compression/decompression, overlay, digital video effects [DVE], mixing).
- Audio algorithms (i.e., mixing, fading, equalizing).
- External video equipment control (i.e., VTRs, DVE units).
- Interface to the digital storage subsystem.

The host processor is responsible for running the operating system, as well as the application software. Host-based software controls the graphical user interface (GUI) and translates the user operations into system commands for the video and audio adapters. The Intel Pentium processor is fast enough to provide and control multiple streams of synchronized digital audio and video data to the specialized adapters, field accurately and in real time. The advent of the PCI bus, with its high-performance features, low-cost implementation, cross-platform flexibility, and plug-and-play ease of installation, has also been a real boon to professional video products vendors. In addition, Movie-2 bus, an open-standard, digital audio/video, over-the-top expansion bus, allows system designers to interface various PCI, EISA, and ISA adapters together to build virtually any broadcast digital video application.

Windows Software Architecture

Although Windows was originally designed for 16-bit processors

Revised version of a paper presented at the 137th SMPTE Technical Conference and World Media Expo in New Orleans (paper no. 137-8) on September 6, 1995. An unedited version appeared in *Moving Images: Meeting the Challenges*, SMPTE, 1995. The paper was extensively updated in October 1996 to reflect the evolution of Windows, ActiveMovie, and OpenDML technology. Alain Legault and Janet Matey are with Matrox Electronic Systems, Ltd., Dorval, Que., Canada H9P 2T4. Copyright © 1996 by the Society of Motion Picture and Television Engineers, Inc.

(Windows 3.0) and mixed-mode 16 and 32-bit processors (Windows 3.1), professional users now prefer the 32-bit models (Windows NT and Windows 95) that take full advantage of the Pentium processor.

For the vast majority of the millions of Windows users, the principal advantages are ease of use and wide application availability. For the video editing community, Windows also provides a multimedia architecture which began with Video for Windows (VFW), and continues with ActiveMovie. This architecture incorporates many of the elements required for typical video applications. As we will see, further improvements are necessary to meet the requirements of the entire professional video market. These improvements are currently being addressed by the Open Digital Media (OpenDML) consortium.

The OpenDML consortium, established in early 1995, is a group of over 100 software and hardware vendors dedicated to making Windows the platform of choice for professional video, audio, and film producers. In close collaboration with Microsoft, the first project undertaken by the OpenDML group was the definition of enhancements to make the VFW audio/video interleaved (AVI) file format more useful for professional applications. The enhanced file format specification was released by OpenDML in November 1995 and has been incorporated into Microsoft's ActiveMovie multimedia architecture.

The following sections give a background of VFW, explain the improvements to the AVI file format developed by OpenDML, and address additional enhancements to ActiveMovie proposed by the consortium.

Video for Windows

Microsoft originally specified two layers of multimedia services in Windows — high level and low level — and subsequently released VFW, which extended the available driver set. A diagram of this architecture is shown in Fig. 1.

The high-level services consisted of the media control interface (MCI) series of software drivers. Microsoft defined various MCI device types (VCR, digital video, animation, etc.) and implemented several drivers. Other vendors supplemented this layer with proprietary drivers. The high-level MCI drivers tended to operate on files or devices and respond to high-level commands (play, pause, seek, etc.).

The low-level services consisted of the wave audio (wave In/Out), musical instrument digital interface (MIDI) control (midi In/Out), multimedia timer, multimedia file I/O (mmio), joystick (joy), and auxiliary (aux) audio channel (for volume, bass, and treble) drivers.

VFW included an MCIAVI high-level driver for AVI files, and low-level drivers: VIDCAP for video capture, video compression manager for video compression and decompression, and an audio compression manager for audio compression and effects, as well as some AVI file-handling services.

In addition, Microsoft documented an audio mixer specification, which was a way for an application to query and interface with a custom user

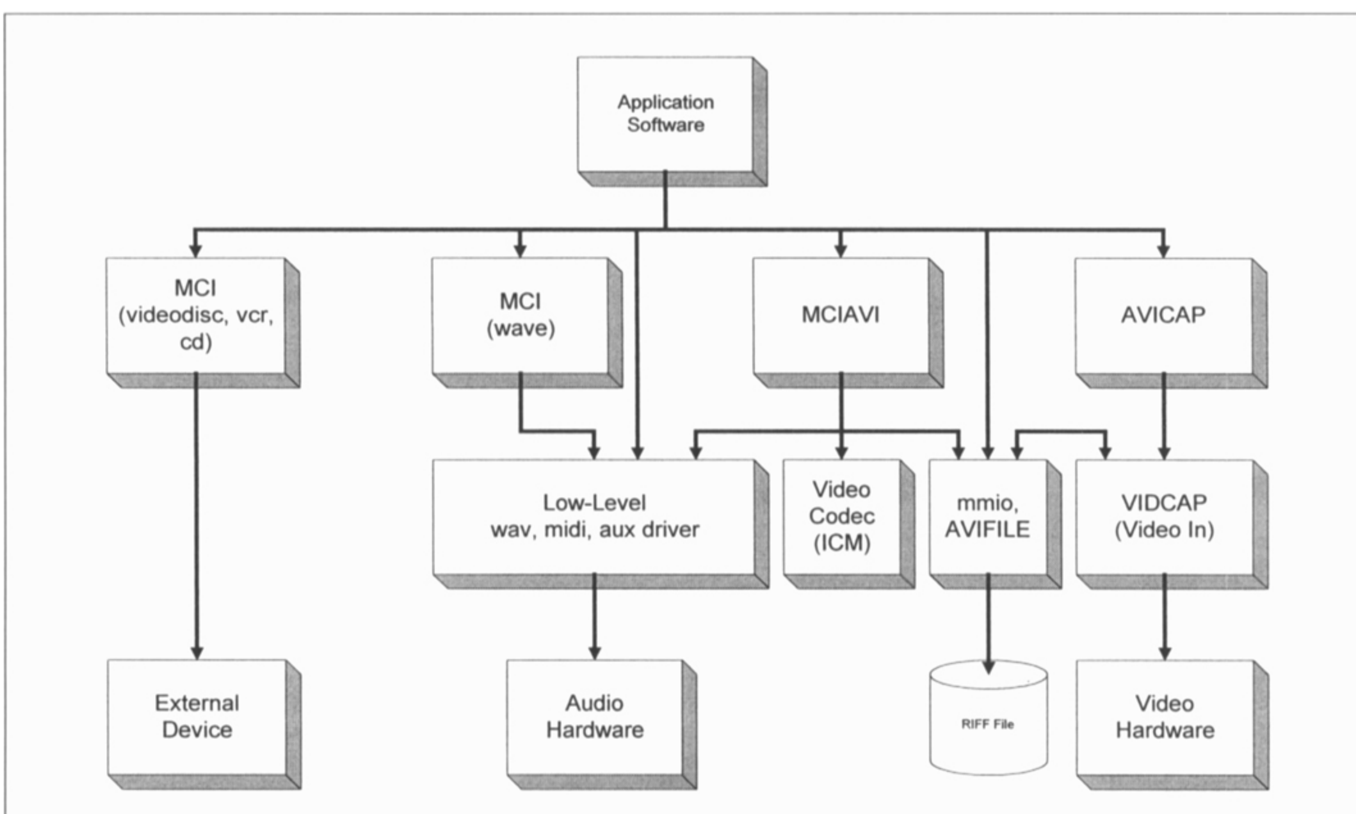

Figure 1. Original Microsoft multimedia software architecture.

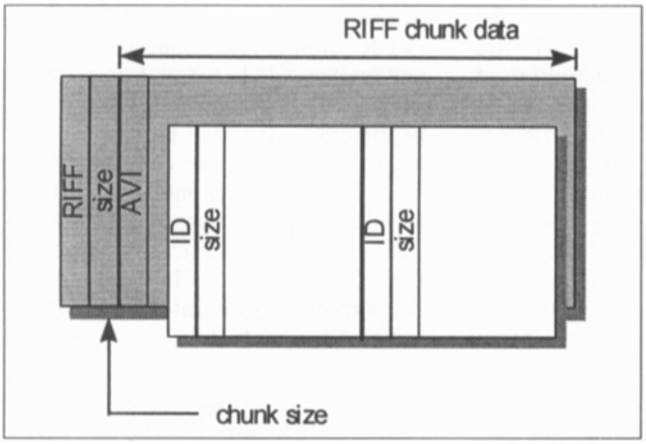

Figure 2. RIFF chunk format.

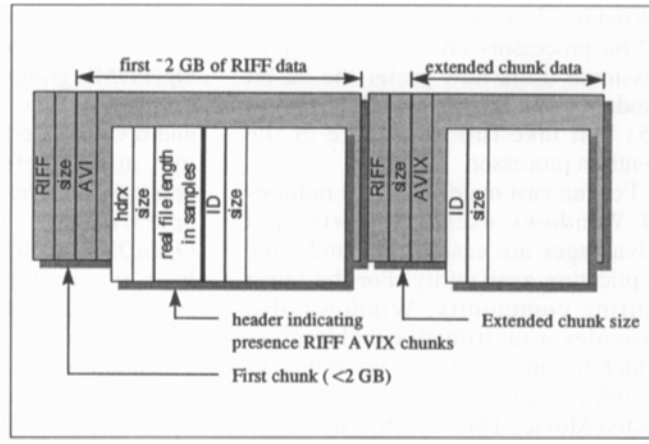

Figure 3. Extension of RIFF format for files > 2 Gbytes.

interface defined inside an audio mixer driver.

Under Windows NT, all of the services were implemented in 32-bit code. The VIDCAP driver application programming interface (API) was not available to the outside world; rather, applications went through AVICAP, Microsoft's API for capturing AVI files. Under Windows 95, some of the services were still implemented in 16-bit code.

Data Types

Microsoft Windows software defines three basic types of data for multimedia: image device independent bitmaps (DIBs), wave audio sample data, and MIDI control data.

Image DIBs

Image DIBs are used to represent static graphic images or single frames inside a motion video sequence. Most computers traditionally have used 24-bit RGB images. In addition to the RGB information, video applications require a key signal associated with the graphics bitmap. This signal, often called an alpha channel, is represented with an 8-bit bitmap. OpenDML proposed to enhance the DIB definitions to incorporate 32-bit DIB structures for 24-bit graphics plus an 8-bit key. Enhanced AVI supports the alpha channel by embedding two streams — one video and one alpha — in the AVI file.

Wave Audio Data

Wave audio data is used by Windows multimedia components to capture, store, and play back digitized audio samples. The typical format is pulse-coded modulation (PCM), and can be sampled at different sampling rates and sample sizes, for example 16-bit, 48-kHz DAT audio.

MIDI Control Data

MIDI is used to control such audio-related equipment as sound synthesizers and audio mixers. Many MIDI devices can be interfaced together in a daisy chain.

OpenDML proposes new data format types for audio and video control data. Specialized audio and video adapters are controlled by the host processor. The adapters need to receive specific field-accurate control data in order to perform user-specified effects in real time. The control data streams include video keyframe information to position video effects and attributes with respect to time. Since video effects adapters are implemented differently by various vendors, OpenDML does not define the specific data but the method and the container of the data structure itself. The data packets are stamped with SMPTE video time code in order to ensure relative video synchrony with the rest of the system.

AVI File Format

Digital nonlinear video editing applications require storage of digital audio and video data streams onto mass storage devices such as hard drives. The file format for such storage must be designed to allow real-time storage, retrieval, and playback. Each video field image data packet inside a digital video stream is typically compressed using Motion-JPEG. The data and format of the surrounding information structures are standardized in Microsoft Windows.

The VFW file format, AVI, is very popular for multimedia applications that play back digital video steams at 320 x 240 resolution at 30 fields/sec. The AVI file format is a resource interchange file format (RIFF) file type. It is composed of data chunks and subchunks, as shown in Fig. 2, which illustrates a RIFF file with two subchunks.

The VFW AVI file format was good as a starting point for OpenDML. The file format was extended to accommodate the needs of the professional video marketplace for high-quality digital video playback at ITU-R 601 resolution and random editing of the sequences in real time. The OpenDML extensions address the issues of interoperability, increased file size, and field versus frame access.

Interoperability

The video data in an AVI file is stored as a DIB or compressed DIB. For video editing applications, the data is usually a Motion-JPEG compressed DIB. Although JPEG is an industry standard, various JPEG hardware codecs from different manufacturers did not interoperate, preventing users from exchanging digital video streams between editing systems. The JPEG bitstream itself is standard, but different chipset implementations of the control elements (code markers) and

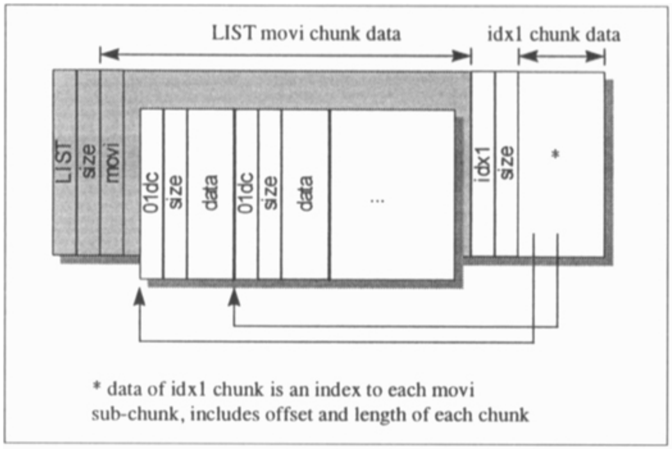

Figure 4. Format of the "idx1" AVI chunk.

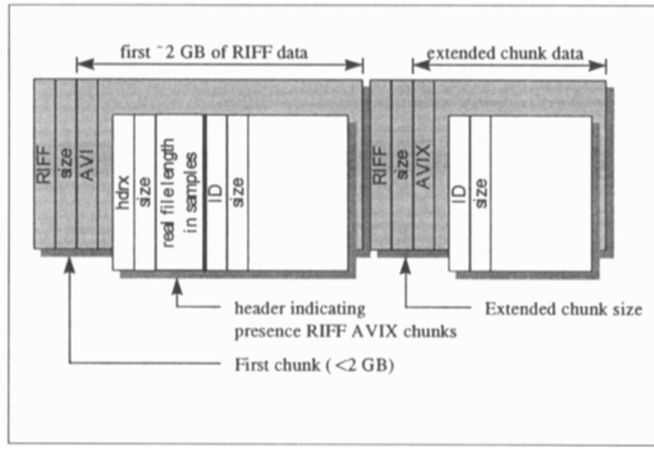

Figure 5. Field index list extensions to AVI files.

byte ordering lead to incompatibility.

This has become a big problem because of the multiplicity of vendors and the availability of digital video networks that would allow the integration of heterogeneous editing environments on a single video server through local area networks (LANs) or wide area networks (WANs). OpenDML addresses this problem by standardizing around the ISO-compliant JPEG bitstream (ISO Standard 10918).

Increased File Size

The maximum file size supported by VFW is 2 Gbytes, which limits high-quality video streams to less than 10 min/file. OpenDML additions to the RIFF format allow file sizes greater than 2 Gbytes by adding extra RIFF chunks to the file, creating a file that can be larger than the maximum RIFF size (Fig. 3).

Field vs. Frame Access

AVI files store each frame of video as an indivisible chunk. Professional video applications require access to the individual fields of a frame in order to render effects such as dynamic motion control (DMC). To accommodate this need, OpenDML extensions to the file allow access to the individual fields.

The AVI file format defines an index list chunk (called idx1) at the end of the file. This list indexes the location of each frame of video stored inside the file, which is necessary for nonlinear applications to randomly reference any point inside the video segment (Fig. 4).

OpenDML defines a set of improvements to the standard AVI index list to ease incremental growth of files, minimize disk seeks to optimize higher video playback throughput, and allow field indexing instead of frame indexing.

The field index list is hierarchical. Field indices are segmented and located physically across the video file. Each individual field index segment (FIS) references a specific block of digital video data contiguous to it. All of the segments are then referenced from a super field index (SFI) list stored at the beginning of the file (Fig. 5).

The field indexing structure has also been enhanced to allow reference to data inside a file using 64-bit instead of 32-bit data pointers. This will accommodate the larger file size described earlier.

Windows File Systems

Microsoft Windows operating systems allow storage of data onto mass storage devices using the FAT, HPFS, and NTFS file systems. Windows 3.1 and Windows 95 support only the FAT file system, while Windows NT supports all of them.

The file system requirements for digital video and multimedia are best handled by NTFS. Multimedia I/O transactions tend to be large in size, limited in number, and nonrepetitive. A sustained flow of data without data disruption is needed to ensure steady playback of video and audio streams. Working with professional video systems developers, Microsoft recognized the need to provide support for these unique requirements. NTFS in Windows NT release 4.0 provides the following capabilities.

Software Striping

Under NTFS, multiple individual hard drives can be configured in parallel to provide overall throughput that is the sum of the throughput of the individual drives minus some overhead. The digital video data stream is split by the NTFS software and chunks of data are written in a round-robin fashion to each disk in turn. For example, a 12 Mbyte/sec system bandwidth can be achieved using three hard drives that each deliver data at 5 Mbytes/sec.

Optimized High Data Throughput

While the data packet size limit for read and write operations under FAT is 64 kbytes, NTFS supports much larger I/O data packets requests. In Matrox DigiSuite, for example, 1-Mbyte transfers are used. This optimizes system throughput.

Virtually Unlimited File Size and File Quantity

Typical professional video products require a large number of large files. NTFS allows for a great number of files to reside on the storage devices and each of these files can be extremely large. Conventional file systems limit file size to 2 Gbytes, which would limit a high-quality video file to approximately 10 min.

Ability to Queue Disk I/O Requests

Typically, a professional system

will play multiple digital audio and video streams simultaneously. NTFS allows the system to anticipate the upcoming data packets required from the disk subsystem. To optimize the accesses to the disk subsystem, NTFS queues I/O requests in advance. While reading the current data packet, the I/O subsystem queues the request for the next data packet.

Reduced Disk Fragmentation through Data Storage Contiguity

To further optimize disk performance, the file system must promote contiguity and prevent excessive fragmentation of the digital video and audio information. The smallest accessible storage unit on a physical disk is called a "cluster," which is a collection of contiguous disk sectors. NTFS features a cluster allocation algorithm that promotes usage of the larger contiguous blocks of free clusters inside the file partition while writing to disk. NTFS contains an image of the free cluster region and dynamically uses it to allocate free clusters to the disk writing process. This eliminates lengthy defragmentation operations as the disks get full.

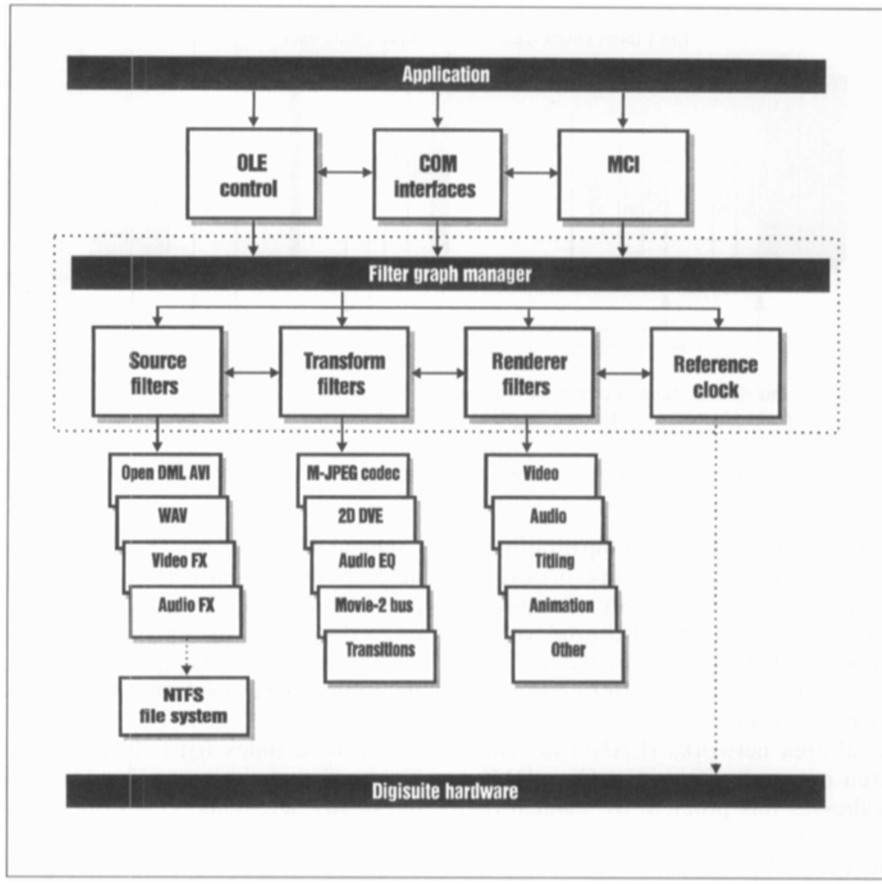

Figure 6. ActiveMovie software architecture as implemented in Matrox DigiSuite.

Other Optimization Factors

NTFS also supports the ability to read and/or write multiple files concurrently, allowing professional video applications to control multiple streams of audio and video simultaneously.

ActiveMovie Architecture

ActiveMovie is a multimedia streaming architecture for Windows NT and 95 that delivers high-quality video playback, while exposing a flexible, extensible, and future-proof set of interfaces upon which digital video applications can be built. It replaces multimedia streaming services and APIs provided by Microsoft in previous versions of the Windows software development kits (SDK) such as VFW and MCI. The basic ActiveMovie software architecture is shown in Fig. 6.

ActiveMovie overcomes all VFW limitations. The improved audio/video file format that was originally proposed by OpenDML has been incorporated in ActiveMovie. High-performance I/O throughput is ensured by the new data streaming architecture. Driver models now have a consistent Application Program Interface (API). They support more data types and are modular, leading to a high level of software/hardware interoperability. System-level, accurate synchronization for all media types is provided and ActiveMovie API is identical for Windows 95 and Windows NT, minimizing the work required for software developers to support both operating systems.

ActiveMovie Data Types

ActiveMovie allows the system designer to define a wider variety of data types than was possible under VFW. For example, in the Matrox DigiSuite implementation of ActiveMovie, the following multimedia data types are defined:

• Digital video data from various types of codecs — Motion-JPEG.
• Digital audio data — PCM WAV.
• Digital video F/X control attributes — 2-D or 3-D DVE attributes.
• Digital audio F/X control attributes — EQ, mixing, and panning.
• Graphics and still images — 32-bit RGB-Alpha images.
• Scroll and crawl of graphics overlays and titles — inscriber scroll and crawl files.
• VCR control — RS-422 Sony Betacam protocol control.

The ActiveMovie data streaming architecture allows the optimal, timely interchange of all these data types. There are three major benefits to the data streaming approach: data throughput is maximized by the ability to use large data buffers (more than 64 kbytes); CPU-intensive memory copying operations, which were required under VFW, are eliminated by the use of shared memory buffers; and system-level synchronization is ensured through the use of time stamps on all data streams.

ActiveMovie Filters

A filter is an object that performs a single task in a multimedia system.

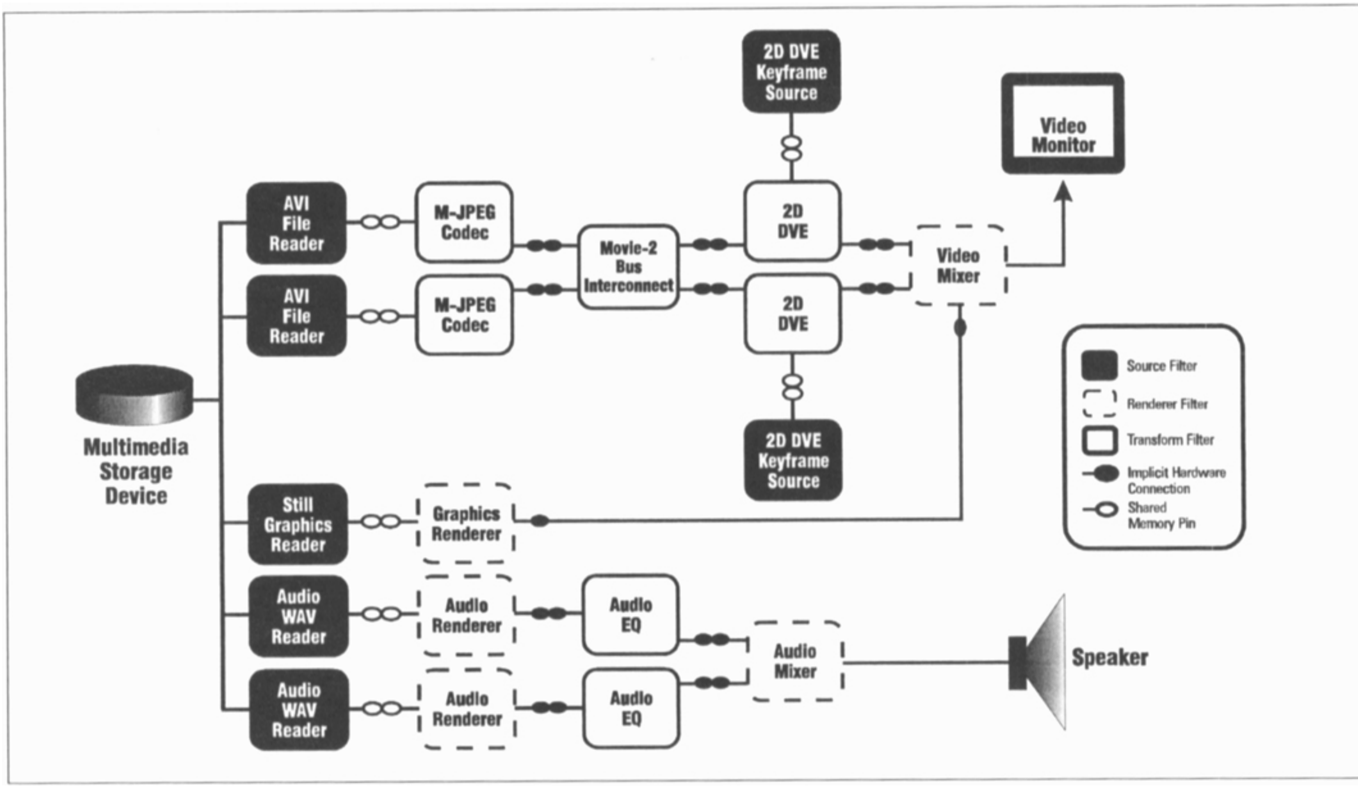

Figure 7. Simplified example of one DigiSuite filter graph.

For example, in a complex subsystem such as the Matrox DigiMix digital video mixer, individual filters perform such tasks as background and wipe generation, chroma keying, layer priority selection, proc amp adjustments, 2-D DVE, etc.

More complex functions can be performed by interfacing multiple filters together to act on a specific data type. Filters are connected by interfaces called "pins" in ActiveMovie terminology. Different types of ActiveMovie filters can either provide, transform, or consume data. These three functions are accomplished using source filters, transform filters, and renderer filters.

Source Filters

A source filter provides data such as digital audio, digital video, or graphics to other filters downstream. Alternatively, it can provide control information such as video and audio keyframes. For example, a data source filter would typically perform functions such as an AVI file reader, WAV audio file reader, and title animator. A source filter can get its control information from a file or from interaction with the user via a user interface device like a scroll bar or fader control. A source filter typically has only an output pin.

Transform Filters

A transform filter accepts a data stream at its input pin, performs a transformation on the data, and provides the processed data to its output pin. For example, in Matrox DigiSuite, the video codec, the 2-D DVE processor, the audio equalizer, and the Movie-2 bus interconnect are just a few of the many hardware-assisted transform filters.

Renderer Filters

A renderer filter is responsible for consuming the processed data and relaying it to a presentation device such as a video display or a speaker. The renderer will ensure presentation of each media stream at the correct time, based on the system-level synchronization mechanism. In Matrox DigiSuite, for example, the "correct time" is defined as presentation of each media stream accurately synchronized at the video field level. Typically, a renderer filter has only an input pin.

ActiveMovie Pins

Two types of pins are defined in ActiveMovie: shared memory buffers and implicit hardware connections. Shared memory buffer pins are used whenever a filter interfaces to another through the use of computer memory. Hardware connections are made when a filter employs hardware acceleration. Because pins are standardized for a given data type, a shared memory buffer pin can be easily replaced by a hardware connection without affecting the interface with the application software. This transparency between hardware and software allows the same application to easily migrate from software-only operation to a higher performance level using hardware accelerators.

ActiveMovie Filter Graphs

A multimedia system requires the connection of multiple filters to accomplish system-level functions. A representation of the multimedia system is called a filter graph (Fig. 7). Filter graphs are customized connections performed between many filters in order to accomplish a specific task.

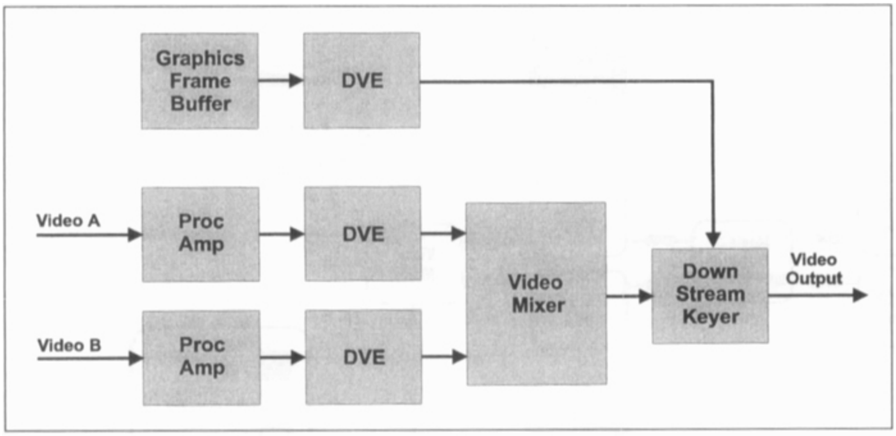

Figure 8. OpenDML video switcher model — A/B + Graphics.

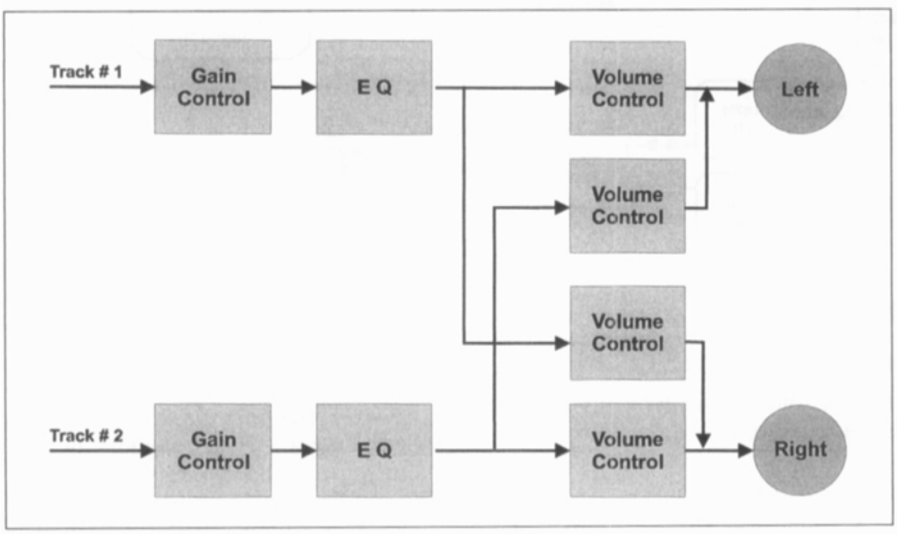

Figure 9. OpenDML audio mixer model.

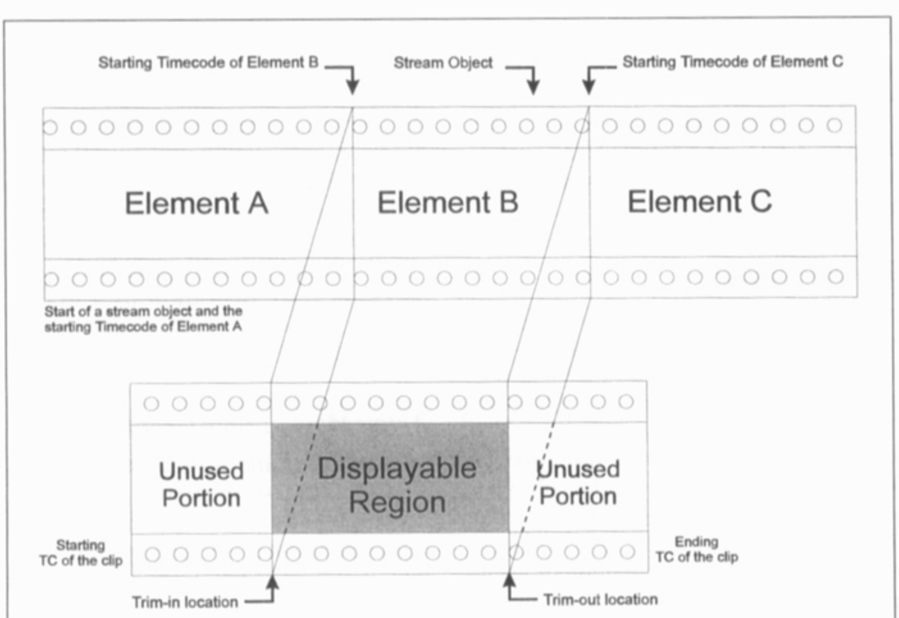

Figure 10. Video stream cut list.

In order to establish a connection between the input pin of one filter and the output pin of another filter, a data-type negotiation takes place. A connection can be established between two filters that share the same data type (i.e., Motion-JPEG digital video) but cannot be successfully established between two inconsistent data types. This prevents feeding a digital video stream into an audio equalizer, for example.

ActiveMovie System-Level Synchronization

The goal of system level synchronization is to present all streams of data to the presentation device in a timely manner. The system designer must decide what the time base or reference clock will be for his application. For example, in Matrox DigiSuite, the time base is the video field (60/sec NTSC, 50/sec PAL), so the reference clock is derived from the hardware vertical synchronization signal provided by the DigiSuite video card. In the film industry, a system designer would likely choose a reference clock of 24 frames/sec. The rendering filters for all the data streams in the multimedia system must be controlled by this single reference clock.

OpenDML Software Extensions to ActiveMovie

ActiveMovie 1.0 is primarily designed for MPEG playback and Internet applications. Because of its extensible nature, however, it is also an important enabling technology for high-performance professional applications. Some developers have already demonstrated that ActiveMovie provides real-time multistream capture and playback of digital video, audio, video special effects (F/X), audio F/X, graphics, and animations in a field-accurate and synchronized manner. This is the functionality required in professional video applications. To standardize various vendors' approaches to these functions, the OpenDML consortium is working with Microsoft to define the features of ActiveMovie 2.0 and develop standard APIs for extensions needed in the video industry.

The first project undertaken by the OpenDML group was the definition of enhancements to make the Video for Windows AVI file format more useful for professional audio/video/film applications. The enhanced file format specification was released by OpenDML in November 1995 and has been incorporated by Microsoft into ActiveMovie. The new file format defined improvements to VFW for greater interoperability, maximum throughput, larger file sizes, lower overhead, and greater editing and random-access capability.

To maximize its usefulness in professional applications, ActiveMovie requires additional extensions. In early 1996, the active members of OpenDML established a Device Driver Workgroup that has taken on the task of identifying, studying, and recommending these extensions to the rest of the industry to ensure full interoperability between vendors, one of the major benefits promised by the ActiveMovie architecture. The extensions will also significantly ease the process for software developers to support multiple hardware platforms.

The general goal of the Device Driver Workgroup is to standardize the various media types needed for professional video applications: digital audio, digital video, video effects descriptors (2-D DVE, 3-D DVE, transitions, chroma key, luma key, alpha key); audio effects descriptors (volume control, EQ, pan, mix, soft cuts, reverb, echo, etc.); animation; and graphics. The Workgroup is also identifying a standard API for various filters: DVEs, mixers, codecs, etc. Since ActiveMovie 1.0 is a playback tool, a capture model is also needed.

Finally, the Workgroup is identifying a standard common functional subsystem — a "filter graph" in ActiveMovie terminology — that will allow software developers to program application software independently from hardware functionality. Common filter graphs for audio and video subsystems are being defined. The initial proposal for the video subsystem model is an A/B roll switcher with graphics downstream keying, DVE and proc amp control (Fig. 8). The audio mixer subsystem features a variable number of tracks with pan, mix, and EQ (Fig. 9).

One important use of the extensions to ActiveMovie is the seamless playback of cut lists with effects in real time on all types of data streams (Fig. 10). This is particularly important for nonlinear editing applications. In real-time NLE applications, the enhanced AVI file format provides fast access to individual frames within an AVI file and it also allows the definition of input and output trim points. ActiveMovie provides the streaming architecture and support for all media types — audio, video, video F/X, audio F/X, graphics, and animations.

Conclusion

Windows 32-bit operating systems have evolved to the point where they can provide the software platform professional video systems designers need to support multiple streams of video, audio, graphics, effects, and other multimedia streams in real time. The key elements that make this high level of performance possible are the enhanced AVI file format that was developed by the OpenDML consortium, the Windows NT file system (NTFS), and the ActiveMovie streaming architecture.

The next step in the evolution will be extensions to ActiveMovie for maximum performance in professional video applications and the standardization of media types, filter APIs, and filter graphs to ensure full interoperability among various vendors' products. The OpenDML consortium, in close cooperation with Microsoft, is addressing all of these issues.

The ultimate result will be the realization of the "open platform" promise of the Windows operating system, with many price/performance choices available to end users in the professional video marketplace.

Appendix — Suggested Readings

Additional information about ActiveMovie in the professional video arena can be found on the World Wide Web at http://www.matrox.com/videoweb/actmvwh.htm.

Additional information about PCI-bus can be found at http://www.pcisig.com.

Additional information about Movie-2 bus can be found at http://www.matrox.com/video.

Additional information about Windows NT can be found at http://www.microsoft.com/NTWorkstation.

THE AUTHORS

Alain Legault

Janet Matey

Alain Legault is vice-president, product development, for the Matrox Video Products Group, Matrox Electronic Systems, Ltd. He is responsible for all aspects of video technology evolution at Matrox including product definition, hardware design, software architecture, and technical marketing. Legault holds a B.E.E. degree from Ecole Polytechnique de Montréal. He is the founder of the OpenDML consortium and the leading proponent of the open-standard Movie-2 expansion bus. He is a frequent lecturer in electronics and computer-aided engineering (CAE) at both Ecole Polytechnique and McGill University in Montreal.

Janet Matey is marketing communications director for Matrox Video Products Group. She is responsible for all aspects of marketing communications at Matrox, including product literature design, advertising, press relations, trade show organization, and telemarketing. Ms. Matey holds a Bachelor of Commerce degree and an M.B.A. from McGill University. She is a regular contributor to video and computer industry publications.