The Digital Compression Facility - A Solution to Today's Compression Needs

Robert C. Plotkin

IBM Visualization Systems, Thomas J. Watson Research Center
P.O. Box 704, Yorktown Heights, NY 10598

Abstract

Compression of motion video and hi-fidelity audio for real-time playback targeted at a number of different playback platforms has entered the domain of big buisness. The choice of architectures for this compression is varied. The most versatile solution for compression uses a general-purpose supercomputer-class computer. This choice allows for the highest quality MPEG compression by providing the mechanisms for processing of video and audio to follow any number of tunable, user-definable, customizable and computationally expensive operations.

1 Introduction

The veritable explosion of multimedia and related fields in conjunction with the overwhelming increase in motion-video-based information today has forced the evolution of video compression from the development lab into full-featured products found in production houses and service bureaus. Today's communication infrastructures, from local area networks to telephone lines, can be used to distribute digital information to the home or office. Although there are a number of solutions for compression, the use of a general-purpose supercomputer with real-time video capabilities offers the compression community a distinct advantage in the race to generate the highest quality compressed data. A software-based approach to the challenge of creating a digital compression facility (DCF) provides the end user with results that are both superior to hardware solutions, and more flexible. The business needs of today exalt the DCF from a research project to a production operation required to fulfill clients' desires for compression targeted at a number of entertainment and related applications.

The solutions for video compression cover a spectrum of output qualities. The lowest quality levels of video compression are suitable for applications using video for reference or informational viewing. This category, which spans many fields from video-telephony to interactive-learning, is exemplified by the video-phone and other inter-office communications solutions being offered today. This type of compression simply attempts to provide moving images to a target station in some pseudo-real-time format. While high video quality is always desirable, simple existence of video images is satisfactory for the purpose of expressing ideas from person to person, or from information source to person. Compression quality is sacrificed for speed of compression, cost effectiveness and ease of video delivery.

The future of compression strives for archival-quality compression. This no-loss compression would be used for long term storage of video information, providing the user with instant access and editing features. For instance, news agencies would like instant access to any news story or video clip. Editing these clips is also desired. The solution for this problem is not available in a cost-effective method today. However, the methods described in this paper are applicable to this level of compression when it becomes viable.

The DCF solution provides for a combination of the best available compression techniques to create what can be termed entertainment-quality material. This is a grade of video quality in which the public is willing to invest; i.e., people will not pay money for poor quality images. The de-facto standard for acceptance is "VHS quality", as people rent movies on VHS tapes, they would expect any new video distribution methods to match or better this degree of quality.

The classic chicken and egg paradox is true for the challenge of generating entertainment-quality compressed video with today's technology. Until recently, one could not distribute compressed material for entertainment purposes (i.e. for money) because there was no cost-effective way to create material that is of

1063-6390/94 $3.00

an acceptable quality level, and no one has created a platform for general-purpose entertainment-quality video compression because there has not been a market for such development. The business now exists for entertainment-quality video compression, and the solutions for cost effectively generating this material are scarce. The emerging markets of video servers, video-on-demand applications and home entertainment systems using compressed material has dictated what will be industry standard acceptable compression methods.

2 The Compression Standard

The MPEG-1 (Moving Pictures Expert Group) compression specification is an international standard that describes how to decode a bitstream and display the video images and sound. MPEG has proven to satisfy the requirements of entertainment-quality compression. In one unofficial test, video from a new VHS tape was compared side by side with an MPEG compressed version of the same material. The MPEG stream was played back on an off-the-shelf product from Phillips called CD-interactive or CD-I. The majority of the discriminating Hollywood viewers chose the MPEG images over the VHS playback.

CD-I is only one decoding solution that utilizes the MPEG standard. The emerging 3D-O technology also relies on MPEG for its full motion video. It should be pointed out that over an hour of entertainment-quality video can be stored on one CD, either for the above mentioned technologies, or for PC or workstations with MPEG decoding add-on hardware. This 200:1 compression can also be applied to numerous other video compression needs. With an aggregate video and audio bit-rate of less than 1.5 million bits per second (Mb/sec), an MPEG stream can be distributed over standard telephone equipment or local area networks.

MPEG compression differs from more conventional compression techniques in that it is based on motion. The JPEG (Joint Photographers Expert Group) standard is a popular standard for compression, but does not take advantage of the similarities between consecutive images in motion pictures. While JPEG is useful for compressing still images, the quantity of data required and the resulting total bit-rate are too high for any lengthy motion video sequences.

MPEG compresses motion video images in three stages to achieve its extremely high quality at very low bit-rates. There are frames of video that are compressed in a similar way to JPEG images. These intra-coded, or I-frames, are reference frames in a MPEG sequence. They can be decoded with no other video information in the sequence. Forward predicted frames, or P-frames, use information from these I-frames or previous P-frames to code an image. The last type of frame, the B-frame, is a bi-directionally coded frame. It uses information from future and previous I- and P-frames.

While MPEG compression is based on a number of novel techniques such as the Discrete Cosine Transform, it is important to note that the MPEG standard does not describe the implementation of the algorithm for compression, but merely the syntax of the compressed bitstream and the methods for decoding. Thus, the quality of the output fundamentally depends on the algorithm used to generate it. A highly efficient compression algorithm, such as the one developed at the IBM T.J. Watson Research Center, is able to achieve exemplary video quality by utilizing algorithm implementations that feature adaptive bit allocation and quantization.

There are other compression methods beside MPEG, but a combination of the inferior quality at low bit-rates, and the proprietary nature of the encoding algorithms prohibit world wide acceptance of these non-standards.

3 The Compression Process - The Pathway to Success

The ultimate use for compressed material should be ascertained prior to the commencement of a compression project. If compressed video material is for reference or informational purposes, than the realm of compression solutions is large due to the probable lack of necessity for entertainment-quality viewing. If the goal is to sell the compressed material, implying that impeccable compression quality is required, then the solutions for compression become much more limited. The remainder of this paper describes a method that allows for the highest quality compression for clients or customers where quality is of the utmost concern. After all, the concept of pay-per-view video-on-demand applications are not viable if the quality of the material does not create a demand!

It is not any single step, but the collection of processes in the DCF which yields marketable compressed video. The ultimate success of a compression project is greater than the sum of all of the steps involved. Compression is more than simply applying a good "compression" algorithm to a piece of source material and generating a result. There are a number of operations that can be performed on the video material in the compression cycle of the DCF. A general-purpose supercomputer-class machine provides the needed flexibility and resources to prepare and operate on video material to generate the highest quality compressed data.

Firstly, MPEG streams may be of varying resolutions, typically from 352 wide by 240 high up to 720 wide by 576. Typically the one-quarter resolutions of 352x240 up to 384x352 are used. A high performance video encoder must be able to handle any and all of these resolutions and aspect ratios if it is to be part of a successful product. Different decoders, and thus projects, require these different resolutions. It is wrong to limit the compression engine to a single, or subset of all resolutions. The compression business, after all, is still emerging and may require new resolutions for new business opportunities.

Secondly, all video has not been created equal. No two films or videos are identical. Why should the compression pathway be the same for them? Those skilled in the art of video compression realize the power of pre-processing video material before compression. The fundamental reason for this step is quite simple. When you compress material 200:1, there will be losses from the original material to the compressed bitstream. These compression losses can, however, be perceptibly reduced by the judicious application of various filters to the original material. These filters, which typically also perform the decimation from the source material to the proper compression resolution, can be selected for varying source material. If the video input material is operated on in the form of a computer file, then any number of processing operations can be applied to it yielding the best possible results. A real time hardware based solution may have vast capabilities, but can never match the unbounded capabilities of a general-purpose solution.

Finally, motion video can originate on film. Material that originates on film at 24 frames per sec (fps) is converted to video (25 or 30 fps) for broadcast and distribution. The majority of video-on-demand and point-of-viewing decoding will use material originating on film. The conversion process duplicates fields of the converted film in the video domain.

For compression applications though, material originating on film should be inverse 3:2 processed (putting the original film frames together again) to yield full film frames that can then be scaled down to the required resolution. This step is of incredible importance because it allows MPEG-1, a non-interlaced standard, to use all of the information from the original film frame. The power of a general-purpose computer solution to this already complicated task, is that a long segment of video material that originated on film can be scanned and re-scanned to find the numerous cadence changes of the 3:2 conversion. Sequences of video that are edited are nearly impossible to decipher on the fly. Having video files available for scanning results in more accurate film frame reconstruction, ultimately allowing for higher quality compression results.

4 Audio Processing

Along with the picture information in a video is audio information. Typically, video formats such as D1, a digital component format, contain 48KHz sampled, pulse code modulated, four channel digital audio which must be compressed at the same time the video is. The compression process must be able to select audio channels as well as convert from the 48KHz to 44.1KHz (most decoders require 44.1KHz MPEG audio although the standard allows for a range of audio rates). Signal-to-noise ratios, distortion and compression artifacts as well as flexible control of compression parameters must be available for the complete digital compression solution.

Audio compression is another process that can take advantage of a general-purpose computer. The audio data, before entering this software compression step can, at the users discretion, be pre-processed for any number of reasons. Typically, audio is left untouched, but a general-purpose software-based solution enables a particular compression project to apply custom processing to the audio. Output may be enhanced with some digital sweetening for a new target MPEG decoder. Again, the option is left up to the user, not limited by a hardware box with fixed audio compression options.

5 The Elusive Systems Layer

Combining or multiplexing audio and video streams together in a systems layer is also an important aspect of the compression process. A compression solution would not be complete if it did not address the needs of real-world decoders, and their requirements for specific encoded data formats or rates. MPEG video and audio streams are individual entities that can be woven together in different ways for different decoding requirements. For instance, the same video and audio compressed data streams (files) will require different system multiplexing techniques depending on whether they are to be viewed on a CD-I player or a video-on-demand server. The full service compression solution must have a multiplexing capability as flexible as the rest of the system.

The raw video and audio streams may be valuable assets on their own, but in conjunction with proper multiplexing software, their value can be expanded. For instance, a real-time compression system may create a compressed stream for one target system, say CD-I. An open-architecture, software-based solution can re-use the video and audio streams to re-multiplex the video and audio for video-on-demand or karaoke without requiring re-running the original source video.

6 The Computer for the Digital Compression Facility

The concept of a video file vs. a video tape is presented to typify the flexibility that is achieved in a software-based compression facility. In the world of video, digital storage implies video tape deck like interfaces with solid state, or Winchester disk storage, but no (or very difficult) direct manipulation of video images. The computer world understands motion sequences, but typically can not handle lengthy 'real-time' acquisition. Many computer architectures utilize expensive add-on hardware to archive short durations of full-bandwidth video acquisition (less than 30 seconds). This does not provide an adequate platform for production video processing and therefore is not suitable for a DCF. The requirements for a software-based solution to the digital compression facility imply a class of machine that can handle both the high bandwidth aspects of compression such as video capture as well as the computationally intensive nature of MPEG compression. Machines such as the IBM POWER Visualization System (PVS), which support the input/output needs of compression including real time digital video capture and storage can form the basis for an end-to-end solution to a wide range of compression requirements. Shared memory architectures with parallel computing capabilities offer computationally balanced solutions for the necessary processing steps involved in the compression solution.

The computational platform selected to perform all the steps required for entertainment-quality digital compression will require a number of features and capabilities to be cost effective.

6.1 Digital Tape Control and Capture

The computer for DCF applications must be able to frame accurately control and have the capabilities to capture data from a digital video tape deck. A DCF would most likely rely on D1 component digital material as source material for compression projects. The DCF must be able to manipulate temporal subsections of a large compression project, i.e., the DCF must be able to perform compression on pieces of a project and then capture different pieces because digital masters of movies will be supplied in multiple reels and computer resources will necessitate that smaller than full tape captures can be processed at any one time.

6.2 Data Storage

A DCF computer requires multi-gigabyte, high bandwidth (greater than 30 megabyte / second) storage. Broadcast quality source video typically is provided on D1 video tape, an industry standard component digital format. Real-time D1 acquisition is a must for a high end DCF. This implies data rates of above the 34 or 27 megabyte / second data rate of D1. Multiple SCSI-II or HiPPI type disk storage subsystems provide the necessary bandwidth, provided the controlling computer has the bandwidth to keep transferring data at a sustained rate.

6.3 Processing and I/O

Captured video needs to be accessed and processed to perform the steps required for the creation entertainment-quality video. A software solution will not be successful if there are bottlenecks in the transfer or manipulation of source or interim video data.

MPEG video compression utilizes a concept known as motion estimation in the compression process. Motion estimation can be described as the process of evaluating sequences of video material and creating motion vectors that specify where an area of similarity from one frame of video to another has occurred. This process requires large quantities of data be traversed. A software solution must have the computational horsepower behind it to be able to perform this as well as other computationally intensive functions such as the filtering and aspect ration correction required for DCF applications.

The benefit to a software solution is that regions of traversal can be changed in accordance with desired compression results. The larger the traversal region, the more computation and thus the slower the process will run, but the better the ultimate compression quality. The flexibility of utilizing adaptive compression techniques frees the DCF from hardware limitations. Each of the steps in the DCF may be adjusted in a software-based DCF. What this means is that every stage in DCF operation can be tuned by the operator to achieve unsurpassable compression capabilities. Additionally, an open software approach to a DCF will allow custom steps to be performed on material, producing material with signature quality by a DCF.

6.4 Flexibility and Extensibility

Fundamentally, a software solution implies that upgrade paths are only a floppy disk away. A DCF solution, such as the PVS, that is 100 percent software based may be converted from todays MPEG-1 standard to new standards such as MPEG-2 at no cost due to hardware obsolescence. Additionally, the lack of any special purpose hardware allows for DCF upgrades without any hardware downtime.

7 The PVS Digital Compression Facility

A Digital Compression Facility based on a general-purpose computer will achieve superior compression results. A totally software-based solution can perform all of the processes described in this paper.

A successful compression product must be flexible and extensible. The digital compression facility should be both a turn-key compression solution that can automatically perform preprocessing steps from inverse 3:2 to filtering, and also be an extensible solution, allowing users to create new and innovative filtering techniques for custom results.

Profitable businesses, most notably in the entertainment industry, tend to utilize unique processes that allow them to stand out or capture a niche market. A successful product for a compression facility must allow the facility to create its own proprietary pre-processing techniques if desired, or allow for custom operation of the facility.

The DCF has to offer all of this and to be truly desirable, offer more. Products, such as the PVS based DCF allow each stage in the compression cycle to be used as is, customized, or replaced by the user. Additionally, material from other sources, such as computer-generated images can also be pre-processed for the compression system. The possibilities of the compression market are only limited by the compression users. New interactive video games will require endless quantities of compressed material and the emerging video-on-demand market is constantly defining new standards for compressed video material.

8 Conclusion

Providing the best quality compressed video requires that the source material be processed through a number of steps including inverse 3:2 processing, filtering, decimation, audio sample rate conversion and video and audio compression before a resultant quality sequence can be realized. Simply providing a high performance compression engine on its own will not yield the desired results. Orchestrating all aspects of preprocessing into a symphony of digital compression steps sets the stage for the premier digital compression facility, and software on large compute server class machines makes it possible.