Lossless Video Compression for Archives: Motion JPEG2k and Other Options

Ian Gilmour, National Film and Sound Archive, Australia
R. Justin Dávila, System Architect and Technology Consultant, Media Matters LLC

Executive Summary

Archivists have a short list from which to select a digital archival file format for a modern, 21st Century migration project. Such an archival file should have the requirement of being both lossless to preserve content and also compressed to conserve storage capacity. The leading candidate for such a file format is Motion JPEG 2000.

Two complementary image compression schemes are in widespread use: JPEG and MPEG, along with some other, less standardized options. The original JPEG has been used for many years in compressing full color images. JPEG-2000 uses wavelet compression as adopted for the Digital Cinema Initiative (DCI), and enables true lossless compression. MPEG-4 is being adopted rapidly for transmission of audio, video and other data at lower bit-rates for similar quality compared to previous MPEG standards. MPEG-4 part 10 Advanced Video Coding, also known as ITU H.264 typically achieves file sizes and data rates half the rates used in MPEG-2. For MPEG-4, in two very specific instances, known as PCM-I, and Transform bypass, the MPEG-4 toolset allows for lossless encoding of finite picture regions or macroblocks. However, when taken as a whole, the MPEG-4 tool set is not designed to support lossless encoding of entire picture sequences, with any significant degree of data reduction. JPEG-2000 is a technically better solution for this requirement.

Why Video Compression?

Modern business requirements for capturing, creating, editing and processing moving images employ a wide range of techniques for reducing the amount of data to be stored and transmitted. Significant advances are occurring in the reduction of bit-rates for end-use distribution and consumer applications such as Internet video streaming, the DCI (Digital Cinema Initiative), Blu-Ray DVD, and high-definition TV. Audiovisual content is typically compressed at every point possible during the production and delivery process, since the human eye is notoriously forgiving of compression techniques in the delivery process. Of course, there is an economic driver in keeping the size of files as small as possible, for both transmission and storage.

For archivists, however, an important distinction needs to be maintained between lossy and true lossless encoding and decoding systems. Perhaps the most basic question facing archivists is how to maintain the essence of audiovisual content in the smallest and most manageable package possible. Typically, compression is performed when an input video stream is analyzed and information that is indiscernible to the viewer is discarded, based largely on the human-perceptual qualities of the content. The perception of the video content is the main driver of video compression technologies, allowing for the perception of a complete audiovisual experience, while benefiting from data compression of the data stream to the extent that compression of the video signal can be achieved without an impact on the audiovisual experience for the end user.

Why Lossless Video Compression?

Ideally, any kind of compression process on the audiovisual content should produce a kind of lossless video compression that is imperceptible to the end user of the audiovisual content while simultaneously maintaining the benefits of keeping all of the information of the source and also the benefits of compression during the production process. Lossless compression offers the best of both worlds, with mathematically identical content to an uncompressed file while simultaneously enjoying the benefits of a smaller file size in increased file storage versus uncompressed files, and the benefits of faster transportation of the compressed file.

"Lossless" means that the output from the decompressor is bit-for-bit identical with the original input to the compressor. The decompressed video stream should be completely identical to original. In a non-video context, there exist a number of examples of lossless formats.

Most computer users are familiar with compressed archive file formats such as arc, Pkzip, Alladin's StuffIt, or UNIX TAR files which find redundancy in data or patterns which can be described using less bits. They often reduce the size of documents and spreadsheets by 4- or 5-fold. Typically, the tape drives used in enterprise data storage systems will permit optional data reduction schemes which process arithmetical patterns in source data, into smaller data sets prior to storage. These data sets are encoded and extracted in less than real-time, so that the streaming rates are greater than for writing and reading uncompressed data. These general-purpose algorithms are designed to compress a wide range of business data including office applications, e-mail, databases and binary data. Sony claims a data-reduction ration of up to 2.6:1 with its Adaptive Lossless Data Compression (ALDC) used in AIT data drives, while a ratio of 2:1 is claimed for LTO. While data extraction has to be 100% accurate, these algorithms are not optimised for images and audio data. These different levels of data compression cannot be cascaded for greater effect. Often, a previously compressed AV file [MPEG, for example] will occupy the same or even greater space on the storage device, when general-purpose data-reduction is applied[1].

Wavelet compression is one of the most effective methods of image compression. The algorithm is based on multiresolutional analysis, and this mathematical analysis has been the frontier of lossless compression over the last several decades. Similar to traditional JPEG compression, a wavelet compression algorithm presents an image as sets of real coefficients, though JPEG uses discrete cosines, instead of wavelets. Most of the wavelet coefficients of a typical image are nearly zero, and the image thus is well-approximated with a small number of large wavelet coefficients. The advantage of wavelet compression is that, in contrast to traditional JPEG's discrete cosine algorithms, the wavelet algorithm does not divide image into blocks, but analyzes the whole image. The characteristic of wavelet compression allows to get best compression ratio, while maintaining the quality of the original audiovisual signal.

Truly Lossless Compression

The importance of lossless compression has an incontrovertible value to the archival community. Wavelet compression is one of the most effective methods of image compression. The algorithm is based on multiresolutional analysis. The part of the mathematical analysis has been actively developed during last 15 years.

Like the traditional JPEG compression, wavelet compression algorithm presents an image as sets of real coefficients. Most of the wavelet coefficients of a typical image are nearly zero, and the image thus is well-approximated with a small number of large wavelet coefficients. The advantage of wavelet compression is that, in contrast to JPEG, wavelet algorithm does not divide image into blocks, but analyze the whole image. The characteristic of wavelet compression allows the process to get best compression ratio, while maintaining the quality.

Benefits of Compression in Archives

One important requirement in archives is that data is not corrupted in storage or during migration over long periods of time, because content may be re-used and re-purposed over multiple life-cycles. Traditionally, two options have been proposed:

[1] In trials at NFSA, MPEG files typically required a 10% increase in data when subject to ALDC coding.

1. moving and storing uncompressed files or streams which are absolutely identical to the original, but may require large amounts of bandwidth and storage space;
2. compressing moving image files or bit-streams in order to reduce storage requirements and transmission times

While most compression algorithms or data-reduction schemes involve irreversible losses of visual or audio information, the JPEG-2000 schema enables significant levels of data reduction for storage and transmission of both still- and moving-images, and guarantees completely faithful reconstruction of content after decoding.

Rationale for Compression in Transmission

It is important to bear in mind both reasons that signals are typically compressed: for reduced cost in storage and for reduced cost in transmission.

MPEG-2 broadcast transmissions originally required an average of 6 Mb/s for reasonable quality, whereas the latest MPEG-2 codecs (enCODer/DECoder) can reduce this to around 3Mb/s. With MPEG-4, full screen standard-definition video, close to VHS quality, may run as low as 1Mb/s. Quarter-screen and low-quality video conferencing may go even lower. In comparison, uncompressed standard-definition video requires 270Mb/s, while uncompressed HD requires 540Mb/s to 1.18Gb/s. Typical bit-rates for the latest MPEG-4 extensions (version 3) for high-definition and D-Cinema are in the range of 10-300 Mb/s (Megabits/second). An HD stream at 300Mb/s would have 4:1 compression, which is not achievable with any lossless scheme at present, encoding typical picture sequences. MPEG-4 and related schemas are deployed typically for much higher rates of data reduction ranging from only slightly lossy 10:1 to over 100:1.

Principles of Perceptual Encoding

Systems which are optimised for audio and image compression often deploy specialised algorithms which exploit weaknesses in human perception. Ever since people attempted to capture and replay moving images and sound, the constraints of mechanics and electronics have required economies and adaptation of content to suit the recording and transmission systems.

In an obvious example, film and video cameras typically capture only a limited number of full images over time, typically 18, 24, 25 or roughly 30 frames or still-images per second. When these are screened, and conveyed to the human brain, we process these image sequences as if they were continuous, ignoring the Our ability to gloss over these temporal aberrations, or imperfections in the time-domain, underpin an important set of principles exploited in image encoding.

Similarly video signals normally don't convey the full resolution of color information that we may see in real life. Some video systems used in domestic cameras and computers have only half as much video resolution, typically denoted 4:2:0

For many years, analog television has also "fooled" our perception by capturing and conveying images in a series of lines drawn across the screen of a television receiver or monitor. Rather than presenting a whole frame at a time, the film is projected, the video signal traces a series of lines across the inside face of a phosphor-coated cathode-ray tube [CRT] which glows momentarily in response to the applied voltage. Because the light emitted from the phosphor coating decays rapidly, each frame is divided into halves known as fields, carrying alternate lines which are sent and displayed in succession. This "interlaced" signal reduces the bandwidth needed in electronic processing, and minimises flickering in CRTs.

Color television takes this a step further, with 3 electron guns firing 3 separate Red, Green and Blue traces onto the inner front surface of the CRT, with each color separated by a mask. In digital video, and computer monitors, images are represented by arrays or mosaic of individual squares or rectangles, each with their own RGB or YUV values. Typical sizes are:

640W x 480H [VGA], 800 x 600 [SVGA], 1024 x 768 [XVGA]
720 x 480 [ATSC SDTV], 720 x 576 [PAL-DVB SDTV], ...
[1280 x 720, progressive-scan HD TV], [1920 x 1080, interlaced HD TV].

Although the basic technology has limitations, and is a long way from providing a seamless presentation of images in the horizontal, vertical and temporal dimensions, our visual perception is good at glueing together all these visual cues into a sense of reality. Just as importantly, global standards ensure that all the basic paratmeters are agreed, so that content can be exchanged without loss or corruption. A film which is shot at 24 frames each second, can be projected at the same speed anywhere in the world, even if converted into digital video. An important principle has been adopted within archives around the world; images and audio should not undergo further degradation each time they are stored and recalled for re-use, otherwise the future portrayal of history would eventually disappear.

Intra-frame Compression

The basis for traditional JPEG and MPEG encoding of data within individual frames of video is the application of a Discrete Cosine Transform [DCT] to shape individual blocks of image data into similar patterns or sequences of pixels, followed by entropy encoding, which describes these areas in mathematical short-hand. This **intra-frame** compression saves large amounts of data when compared to the task of storing or transmitting a full description of every pixel in the original image. MPEG-2 and MPEG-4 standards have enabled increasingly adaptive borders of macroblock boundaries to match more closely the transitions in image complexity.

MPEG-4 Part 10/H.264 AVC includes an additional post-filtering stage to smooth the edges of macroblocks which are visible typically at higher compression ratios. JPEG-2000 includes a more advanced intra-frame compression technique known as wavelet, which minimises blocky artefacts when compared to DCT compression.

Inter-frame Compression

MPEG video compression supports a hybrid encoding regime, in which not only individual frames may be compressed, but entire sequences of frames with similarities [usually within a shot or scene] may be encoded with data specifying only the differences between adjacent frames. Under certain conditions, typically at lower bit-rates, these differences over time, known as temporal or inter-frame encoding techniques, require less data for a given picture quality, than does encoding every frame. Inter-frame encoding typically maps groups of pixels within macroblocks which stay the same from one frame to the next [i.e. fixed backgrounds] or which move in the same direction [e.g. moving objects or panned backgrounds]. Rather than recoding these image regions, their relative positions are tracked using motion vectors, which require much less coding.

A typical group of pictures (or “GOP”) would include one I-frame which is intra-coded [so that it can be represented on its own], along with P-frames which are predicted using such tools as motion-prediction and motion-compensation. Additional B-frames (“in-Between”) are then extrapolated from I- and P-frames. A typical long-GOP MPEG structure would be: I-B-B-P-B-B-I-B-B-P-B-B-I-B-B-P-B-B-I-B-B-P-B-B-I for a 25 frame sequence.

Two Complementary Image Compression Schemes

Two main families of image compression are now in widespread use, JPEG and MPEG. The JPEG standard has been developed by the Joint Picture Experts Group [Working Group 1 under Sub-Committee 29 of the ISO/IEC Joint Technical Committee-1]. JPEG sits alongside MPEG [WG-11], under SC-29 of

Joint Technical Committee 1 in the ISO standards framework.[2]

JPEG

The original JPEG [Joint Picture Experts Group] schema from the early 1990s has been extended and enhanced recently with JPEG-2000, which supports wavelet compression, and mathematically lossless data reduction with absolutely no effect upon image integrity. Both JPEG and MPEG families are backwardly compatible, so that older files can be decoded with newer software. Generally, the focus of the JPEG group has been to preserve the value of each frame that is being compressed, which lends itself uniquely to the archival community.

MPEG

The MPEG standards have evolved from the original MPEG-1 audio and video compression schemes in 1991 [check date], into MPEG-2 now used for digital cable, satellite and terrestrial broadcasting, DVD, High-definition and many other applications. Improvements include adaptive macroblock boundaries for better definition of picture region boundaries, ability to handle interlaced video, variable bit-rate encoding [VBR], and multi-channel surround sound.

MPEG-7 and MPEG-21 have been developed also to define areas such as metadata standards for multimedia, interactivity and intellectual property rights.

MPEG standards typically outline a set of parameters which may be used to encode and decode audio and image sequences, and to package this essence along with metadata. The standards specify only those requirements for universal compatibility of decoders, rather than constraining how content is encoded.

The MPEG standards have tended to focus on multi-layered encoding for widespread distribution of content, as opposed to archiving and preserving the original content.

MPEG-4

From Dec 2001, MPEG-4 part 10 was jointly developed by the Joint Video Team from ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) who refer to the standard as H.264. It follows H.263 narrow-band coding for video-conferencing. Related commercial deployments of MPEG-4, AVC include The MPEG-4 standard comprises over 20 parts which are closely related, and which define areas of conformance including visual and audio coding [parts 2, 3, 10]; systems [1]; file format [12, 14, 15], conformance[4]; software and hardware [5, 7, 9]; carriage of content on IP networks [8], IPMP Extensions [13]; delivery multimedia integration framework [6]; scene description and application engine [11], animation framework eXtension (AFX) [16], font compression and text streaming [17, 18]; and MPEG-Java extensions for rendering [21].

MPEG-4 includes several improvements over MPEG-2, including:

- multi-directional motion vectors, ¼-pixel offset;
- object coding for greater efficiency, separate control of individual picture elements;
- behaviour coding for interactive use.

ITU-T Rec. H.264 | ISO/IEC 14496-10 version 1 approved in 2003 categorically states that lossless compression is out of scope[3]:

---

[2] MPEG-4 Industry Forum: http://www.m4if.org/resources/m4-out-30035.pdf

[3] ISO/IEC JTC1/SC29/WG11/N4672 CODING OF MOVING PICTURES AND AUDIO Requirements for AVC Codec: Source Requirements March 2002, Jeju, Korea

Compression Performance

Subjective, visually lossless compression is defined for MPEG-4 as mathematically lossless compression, but is not a requirement for the standard**.**

This original scope statement highlights an important distinction between true, mathematically lossless compression, and lossy compression schemes in which significant amounts of picture information but whose effects are visually difficult to perceive. These systems are also known as “pseudo lossless”. In true, mathematically lossless compression systems, every pixel of every image is transmitted and rendered in exactly the same place within the image frame, with exactly the same value as the original. ITU-T Rec. H.264 | ISO/IEC 14496-10 version 3, approved in July 2005 includes a set of Fidelity Range Extensions [FRExt]which are intended to expand the original application[4].

Other Lossless Video Candidates

In addition to the JPEG group and the MPEG group, there are a number of additional lossless video candidates. In the interest of providing a comprehensive list of all lossless candidates, we will review them all here. None of these should be considered to be the same caliber as JPEG and MPEG in terms of broad and well-published support.

HuffYUV

HuffYUV is a lossless codec that relies on the Huffman code for compression, a popular and memory-efficient code set useful for many compression applications. The HuffYUV video compression algorithm is credited to Ben Rudiak-Gould’s application of this compression algorithm to video.

HuffYUV is intended to replace uncompressed YUV as a video capture format. It is fast enough to compress full-resolution CCIR 601 video (720 x 480 x 30fps) in real time as it's captured on my machine. HuffYUV also supports lossless compression of RGB data.

The criteria for a viable candidate for a long-term preservation video format require that the format be a publicly published format with a reasonable level of commercial support. This is one area where HuffYUV falls short, in that while it is based on well-published it is project supported by a limited staff, namely Ben Rudiak-Gould.

AlparySoft Lossless Video Codec v 1.6.

Similarly, the video codec produced by AlparySoft (www.alparysoft.com) offers a lossless mode, but it does not enjoy the benefit of an international standards board backing this format.

PicVideo Lossless JPEG Codec v 2.10.0.29 Library 2.0.0.400

The PicVideo Lossless JPEG Coec is based on the original discrete cosine compression offered by JPEG. This code is produced by Pegasus (www.jpg.com), and the benchmark configuration used for compression/decompression is 0 low bits discarded and with RGB compression by pseudo lossless YCbCr.

MPEG 4 Revisited: MPEG-4 Lossless is not Truly Lossless

---

[4] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 12th Meeting: Redmond, WA, USA 17-23 July, 2004
Amendment http://ftp3.itu.int/av-arch/jvt-site/2004_07_Redmond/JVT-L047d1.zip

According to MPEG-4 subclause 0.5 "Overview of the design characteristics" of MPEG-4:

> *"With the exception of the transform bypass mode of operation for lossless coding in the 4:4:4/12 profile and the I_PCM mode of operation in all profiles, **the algorithm is not lossless**, as the exact source sample values are typically **not preserved** through the encoding and decoding processes."*

The subclause 7.4.2.1 "Sequence parameter set RBSP semantics" further specifies conditions of transform processing:

> **lossless_qpprime_y_zero_flag** equal to 1 specifies that the transform coefficient decoding process and picture construction process prior to deblocking filter process as specified in subclause 8.5 shall be used when QP'_Y is equal to 0.
>
> **lossless_qpprime_y_zero_flag** equal to 0 specifies that the transform coefficient decoding process and picture construction process prior to deblocking filter process shall not apply the transform bypass mode of operation. When lossless_qpprime_y_zero_flag is not present it shall be inferred to be equal to 0.

As shown in the following extract from Subclause 8.5.6 "Scaling and transformation process for luma DC transform coefficients for Intra_16x16 macroblock type"

– If lossless_qpprime_y_zero_flag is equal to 1 and QP'_Y is equal to 0, the output dcY is derived as

$$dcY_{ij} = c_{ij} \quad \text{with } i, j = 0\ldots3 \qquad (8\text{-}253a)$$

– Otherwise, the following text of this process specifies the output:

After the inverse transform, scaling is performed as follows.

– If QP'_Y is greater than or equal to 36, the scaled result shall be derived as

$$dcY_{ij} = (f_{ij} * LevelScale(QP'_Y \%6,0,0)) << (QP'_Y / 6\text{-}6), \text{ with I, j} = 0..3. \qquad (8\text{-}255)$$

– Otherwise (QP'_Y is less than 36), the scaled result shall be derived as (8-256)

$$dcY_{ij} = (f_{ij} * LevelScale(QP'_Y \%6,0,0)) + 2^{5\ QP'}Y^{/6}) >> (6\text{-}QP'_Y / 6), \text{ with I, j} = 0..3.$$

NOTE – When entropy_coding_mode_flag is equal to 0 and QP'_Y is less than 10 and profile_idc is equal to 66, 77, or 88, the range of values that can be represented for the elements c_{ij} of c is **not sufficient to represent the full range of values** of the elements dcY_{ij} of dcY that could be necessary to form a close approximation of the content of any possible source picture by use of the Intra_16x16 macroblock type.

In new subclauses A.2.4 "4:2:0/10 profile", A.2.5 "4:2:2/8 profile", and A.2.6 "4:2:2/10 profile":

– Sequence parameter sets shall have lossless_qpprime_y_zero_flag equal to 0 only.

– Only I, P, and B slice types may be present.

In other words, the lossy entropy coding transform applies when data reduction is required, or, alternatively, the RBSP semantics can specify that certain picture regions of a frame are either:

1. Passed through without entropy coding [in which case there is no data reduction]; this mode only applies to 4:4:4 profiles, not to 4:2:2 or to 4:2:0;
2. Transformed with coding that will reconstruct specified picture regions within a frame, with a possibility of increased data requirement, and losses incurred in other parts of the process.

This does not constitute an accurate, mathematically lossless solution for start-to-finish image transmission or storage. Rather, it provides options for non-coding where the law of diminishing returns would apply.

Other Lossless Data Reduction Systems

Most computer users are familiar with compressed archive file formats such as arc, Pkzip, Alladin's StuffIt, or UNIX TAR files which find redundancy in data or patterns which can be described using less bits. They often reduce the size of documents and spreadsheets by 4- or 5-fold. Typically, the tape drives used in enterprise data storage systems will permit optional data reduction schemes which process arithmetical patterns in source data, into smaller data sets prior to storage. These data sets are encoded and extracted in less than real-time, so that the streaming rates are greater than for writing and reading uncompressed data. These general-purpose algorithms are designed to compress a wide range of business data including office applications, e-mail, databases and binary data. Sony claims a data-reduction ration of up to 2.6:1 with its Adaptive Lossless Data Compression (ALDC) used in AIT data drives, while a ratio of 2:1 is claimed for LTO. While data extraction has to be 100% accurate, these algorithms are not optimised for images and audio data. These different levels of data compression cannot be cascaded for greater effect. Often, a previously compressed AV file [MPEG, for example] will occupy the same or even greater space on the storage device, when general-purpose data-reduction is applied[5].

Summary

The deployment of a particular encoding and decoding scheme should be based upon sound business principles of choosing the best tools for the job. MPEG-4 is configured typically to achieve significant reductions in data, with some irreversible losses of image and audio quality. MPEG-4 pt 10 or ITU H.264 is a very efficient way of transmitting acceptable quality video at much lower bit-rates than uncompressed bit-streams. This application is clearly optimised for end-user distribution through narrow bandwidths, and where no subsequent re-coding or re-puposing is required. By comparison, JPEG-2000 has concentrated upon the optimisation of image quality within individual frames, and allows true lossless data-reduction for applications such as archiving, where no loss of image quality is acceptable.

About Media Matters

Media Matters, LLC. is a technical consultancy specializing in archival audio and video material. We provide valuable advice, analysis, and advanced technology to improve collection management, preservation and access for major audiovisual archives around the world. www.media-matters.net

About the Australian National Film and Sound Archive

Part of the Australian Film Commission, the National Film and Sound Archive is Australia's audiovisual archive, collecting, preserving, and sharing the nation's moving image and recorded sound heritage. http://www.screensound.gov.au

[5] In trials at NFSA, MPEG files typically required a 10% increase in data when subject to ALDC coding.