Preservation-Worthy Digital Video, or
How to Drive Your Library into Chapter 11

Jerome P. McDonough
Elmer Bobst Library, New York University

Abstract
This paper provides an overview of New York University Libraries' investigations into best practices with regards to archiving digital video, including a brief technical overview of relevant characteristics of digital video, a discussion of the theoretical requirements for preservation-worthy digital video and some discussion of the costs for creating and maintaining a large scale digital video archive.

Introduction

Libraries and archives face increasing pressure to make materials in their collections accessible through digital library systems. While early research on digital libraries has focused primarily on collections of textual and still image material, the growing bandwidth available to both end-users and cultural memory organizations, along with the falling price of disk storage, has meant that the creation of digital libraries of audio/visual materials is increasingly feasible. As a result, many libraries and archives have begun to make parts of their audio/visual collections available online.

As these organizations begin to digitize audio/visual materials, however, some key differences between previous projects, that focused on text and still image works, and new ones focused on time-based materials have emerged. Many early digital library projects focused on materials that are rare, unique, and quite often fragile, but these materials are also to some degree inherently long-lived. Materials like the Tebtunis Papyri digitized by the University of California at Berkeley (Bancroft Library, 2003) are extremely fragile, but they have existed over 2000 years, and with proper care may reasonably be expected to last for some time to come. One of the justifications for digital library projects such as this, in fact, is that digital images in many cases provide an acceptable surrogate for the original, and hence reduce wear and tear due to handling and help preserve the original artifact. They also do not suffer from technological obsolescence; the media will not become unreadable for lack of a suitable player.

The moving image materials that libraries and archives have begun to digitize are, in many cases, proving to be nowhere near as long-lived, both due to the fragility of the underlying media and the rapid rate of technological change in the field of moving images. This has meant that when libraries and archives begin digitizing moving image materials within their collections, they must in many cases consider whether to design

these efforts simply as a means to enhance access to materials in their collections, or whether they should also use such projects as an opportunity (in some cases, a last opportunity) to engage in reformatting essential to preserve at-risk material.

The acceptability of digitization as a preservation strategy is still a matter of some dispute in the library and archival communities. I do not intend to directly address that debate here, although I would note that these debates take on a somewhat different tenor when considering original source material that is already electromagnetic in form, rather than brittle books. Instead, at New York University (NYU) we have been trying to investigate the question of how, if we do wish to exploit digitization as a preservation medium for videotape material, we can best go about doing so. In particular, since digital media offer the promise of lossless copying of information over generations, what is necessary to insure that a digital copy of video material produced today will not decay over time? Answering these questions requires some basic background in the nature of a video signal, and the approaches commonly used to digitize it.

Digital Video 101

A video signal consists of a single luma and two chroma components. Luma (Y') provides the brightness value associated with any point in the video signal; the two chroma components, a red color difference value ($R'-Y'$) and a blue color difference ($B'-Y'$), provide color information for each point independent of the luma value. For those whose experience with video systems has been confined to computer video, where the signal is transmitted as a set of red, green and blue components, this use of luma and chroma may seem peculiar, but historically it has served a valuable function. The use of luma and chroma meant that broadcasters did not need to produce and transmit two different signals to support black and white and color television; a single signal could be transmitted, and black and white sets would use the luma portion of the signal and disregard the chroma information, while color television sets used all three components to generate the viewable image.

The lack of any apparent green component to a standard video signal is also somewhat confusing to those accustomed to computer video color spaces. However, it is possible to reconstruct the value for the green portion of a video display based upon the values of the luma and chroma components.[1] Human beings are somewhat better at perceiving green light than we are at perceiving red light, and better at red light than blue. The value of luma for any point in a video display can therefore be expressed as a weighted sum of the nonlinear red, green and blue components for that point:

$$Y' = 0.299R' + 0.587G' + 0.114B'$$

Given this, the value of green can be expressed as:

---

1. For an excellent discussion of luma/chroma and color space issues with respect to video, see Poynton (2003) and Poynton (1997).

$$G' = \frac{Y' - 0.299R' - 0.114B'}{0.587}$$

If then a video signal provides the value of Y', $B' - Y'$, and $R' - Y'$, it is a simple matter to reconstruct the values of B' and R' by adding the value of Y' to the two color difference signals, and then using the values of Y', B' and R' to calculate the value for G'. In effect, green information is encoded in the video signal as a function of the luma and chroma information.

Sampling of the analog values for the luma and chroma components of a video signal during digitization is typically conducted in a manner rather different from the approach that libraries and archives have used to digitize still image materials. In capturing still images, the usual practice is to sample and record red, blue and green color information for each pixel in the image. The equivalent in video would be to sample and record the luma, red color difference and blue color difference values for each pixel. This is almost never the case in video digitization. More typically, a digitization process for video will record luma for every pixel, but will sample the color difference signals less frequently. Some of the more common sampling regimes for digital video are as follows:

4:2:2 Luma is sampled at every pixel, while the two color difference signals are sampled at every other pixel. This is the standard for most professional digital video equipment.

4:2:0 Luma is sampled at every pixel. Sampling of the color difference signals is alternated every line, with the $R'-Y'$ sampled line for line, then $B'-Y'$ on the next. For both color difference signals, the samples are taken every other pixel. This regime is used MPEG2, and hence is in common use as the standard format for DVDs.

4:1:1 Luma is sampled at every pixel, while the color difference signals are sampled every fourth pixel. This regime is used in DV and DVCAM products.

There are other less common sampling regimes, such as 3:1:0 and 3:1:1 (Poynton 2002).

As with the use of luma and chroma, the use of subsampling regimes in digital video may appear somewhat unusual to those with experience digitizing still image materials. The reason for subsampling in digital video is reasonably simple; without it, the data rates and data storage requirements would present formidable obstacles to implementation. Consider a standard NTSC video frame, with an active display area of 720 x 480 pixels, or a total of 345,600 pixels per frame. If the three channels (luma and two color difference signals) are each sampled at 10 bits (30 bits per pixel), then sampling a single frame will require 10,368,000 bits. There are 29.97 frames per second in NTSC video, so with full three channel sampling at 10 bits per sample, the resulting digital stream would consume 310,728,960 bits per second. Storing an hour of video at this rate would require nearly 140 gigabytes. It is only comparatively recently that computing equipment capable of sustained throughput of 310 megabits/second reached

a price where libraries and archives might consider its use on digitization projects. Relatively few institutions are capable of affording large-scale video storage, if such storage requires 140 GB for a single hour of video. 4:2:2 subsampling drops the data rate and storage requirements for digital video by a third, and the resulting video, when displayed, is almost indistinguishable from the original source. 4:2:0 and 4:1:1 subsampling cut the requirements in half, and still provide perfectly acceptable video quality for consumer applications.

It is relatively common within publications and advertisements aimed at the video production community to see the term "4:2:2 uncompressed" to refer to video which uses 4:2:2 chroma subsampling, but otherwise does not compress the video signal. Strictly speaking, this is an oxymoron. Use of chroma subsampling is a form of lossy compression; color information is being discarded to reduce the size of the video stream, in anticipation of the fact that it can be interpolated relatively successfully from the remaining information when the stream must be displayed to the viewer. Such a reconstructed signal, however, will not be a perfect match for the original.

Chroma subsampling is by no means the only form of compression used to reduce the data rate and storage requirements of video. Standards such as MPEG2 and MJPEG2000 employ a variety of techniques, including motion compensation, discrete cosine transformation and wavelet-based compression to further reduce the size of a video stream to a more manageable level. Current video production practice is built upon a foundation of lossless and lossy compression techniques.

Preservation-Worthy Video vs. Normal Video

While the library and archival world has only limited experience in creating digital libraries of video materials, we have gained a fair amount of experience in digitizing text and still image works. Research and experimentation over the past several years has led us to identify several characteristics a digital file must possess if we are to consider it 'preservation-worthy':

- As formats tend to fall out of use over time, the file must be in a format that will enable us to move its content to new formats without loss of information;
- Similarly, as media inevitably decays, the file must be stored on a medium that allows us to move the information to new media without loss;
- The file must be in a format for which the complete technical specifications are publicly documented (preferably in a formal standard), so that we can examine the format's characteristics to be certain that it will not place information at risk of loss, and so that if necessary we can create new software to access information within the file;
- As users' needs may not be as well satisfied by a digital file we consider preservable as they may be by some other format, to the degree possible, a preservation file should be in a format chosen with an eye towards insuring the easy production of derivative files for distribution to end users;
- Preservation is an expensive activity, and in order to insure the preservation of the largest quantity of material possible, files should be stored in a format that minimizes

our costs of digital production, distribution and migration.

In practice, these requirements have tended to favor standards-based file formats over proprietary ones, as standards insure the availability of the technical information needed to evaluate a particular file format. They also require avoiding the use of lossy compression techniques, as migrating from one form of lossy compression into a new form is almost certain to introduce artifacts into the digital file.

Unfortunately, use of lossy compression is the normal practice in almost every piece of video production equipment employed today. As indicated in the above discussion of sampling, all video processing involving chroma subsampling, even that advertising itself as 'uncompressed', is actually using lossy compression. Because of this, we can anticipate that video streams stored using chroma subsampling (which is to say, the large majority of video in the world at the moment), will be apt to experience artifacting when moved to new formats.

Experiments conducted by Marco Solorio of One River Media (One River Media, 2003) demonstrate the problem. The images below demonstrate what happens to a test image when it is compressed and then decompressed successively using the same 4:2:2 uncompressed codec.
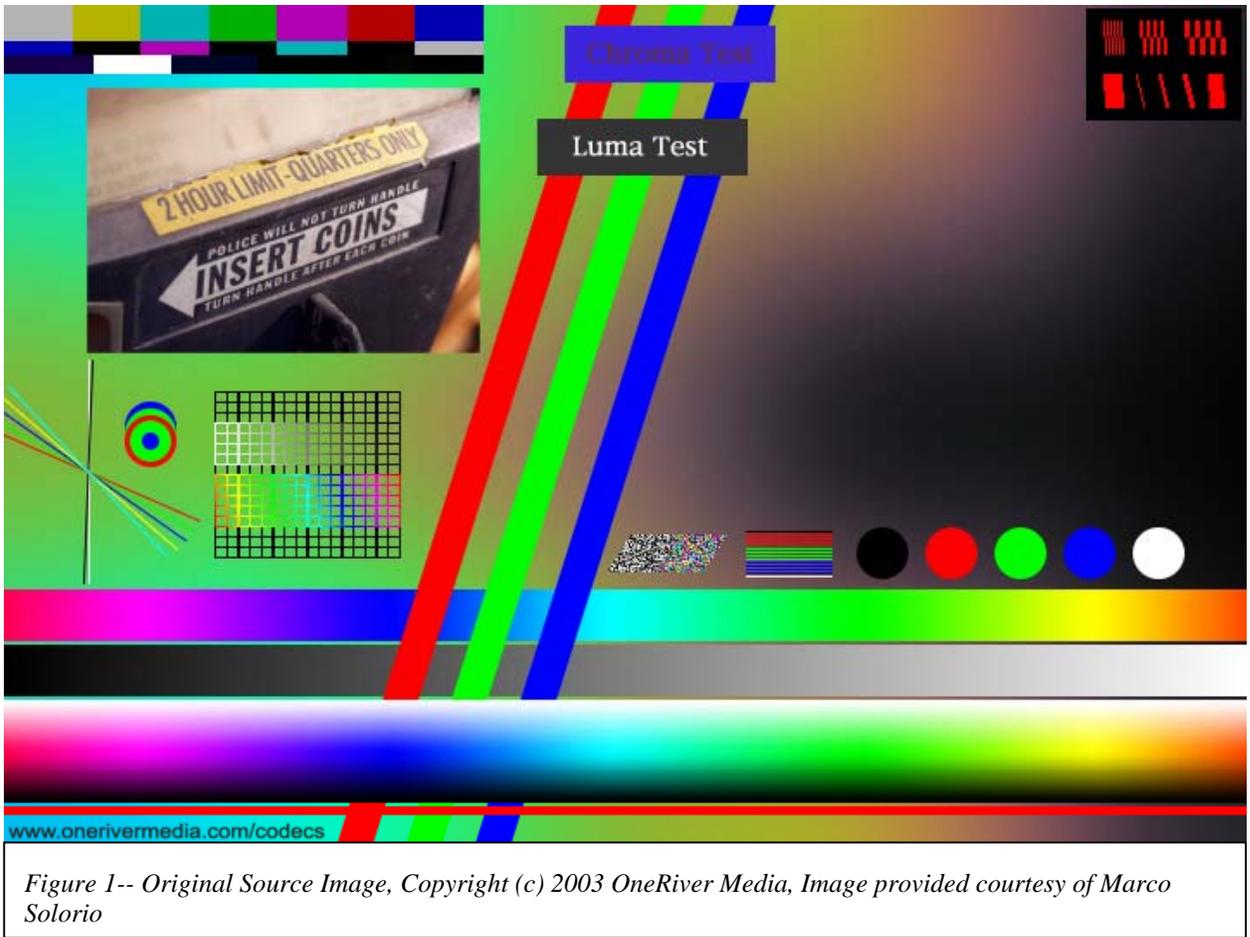
Figure 1-- Original Source Image, Copyright (c) 2003 OneRiver Media, Image provided courtesy of Marco Solorio
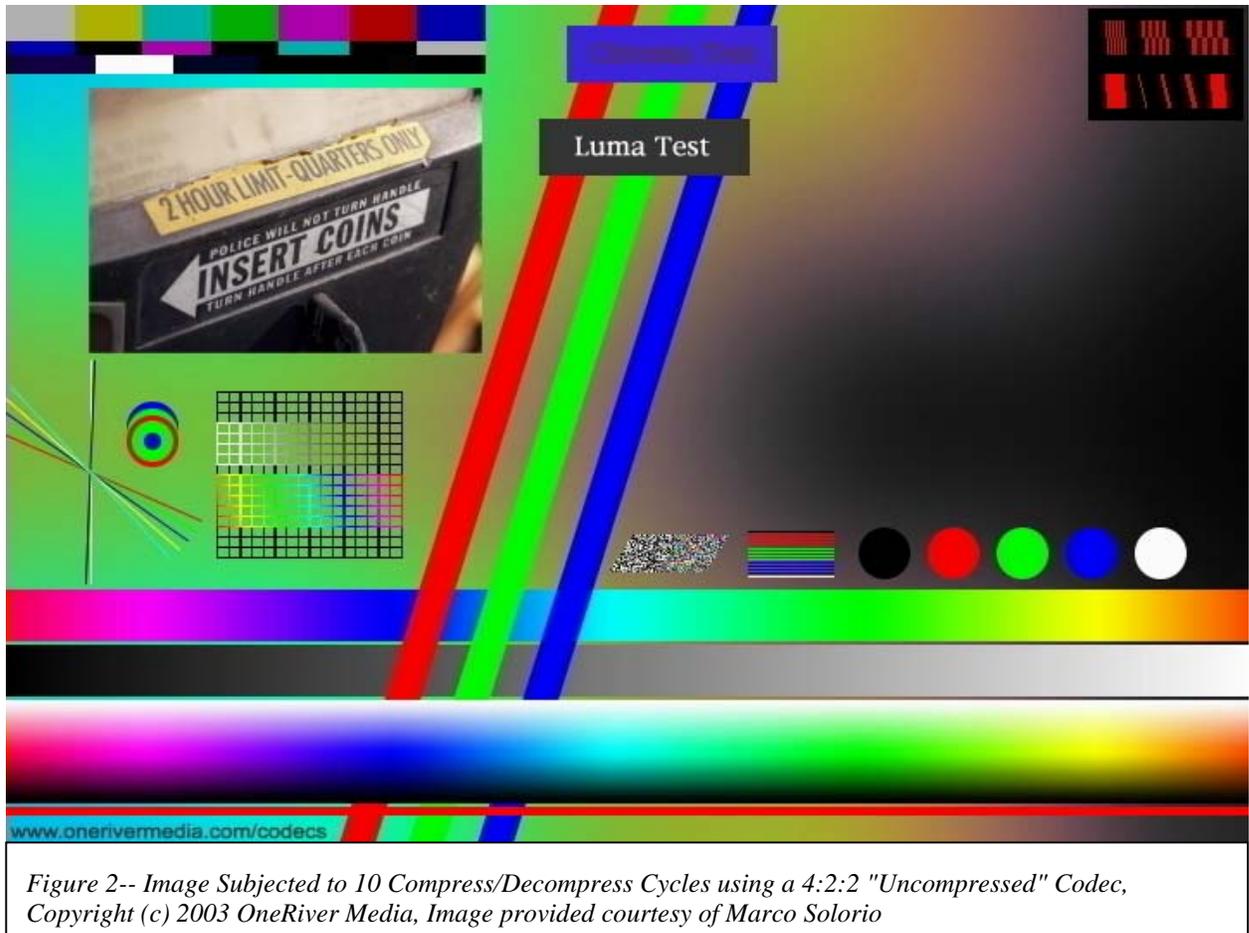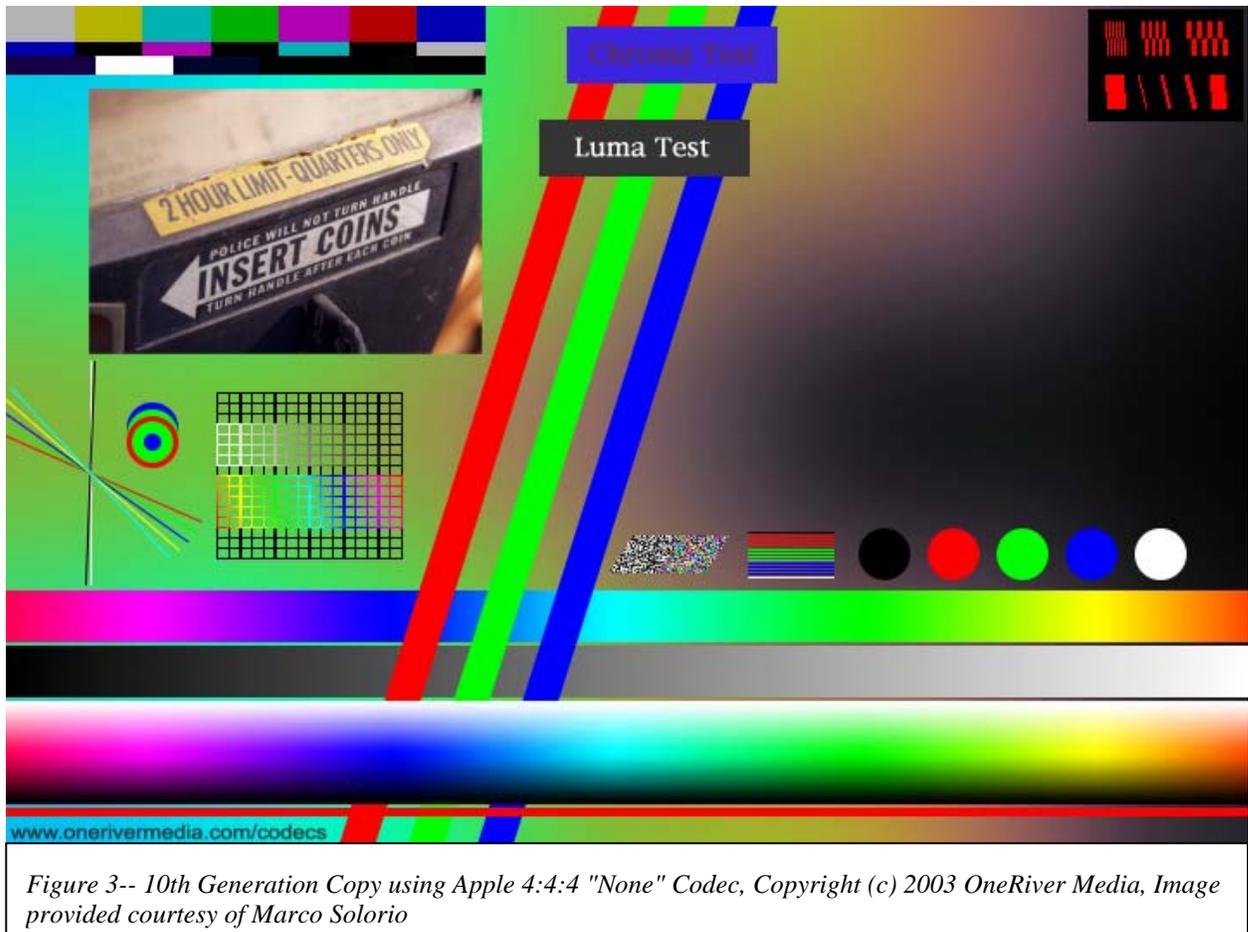
Figure 2-- Image Subjected to 10 Compress/Decompress Cycles using a 4:2:2 "Uncompressed" Codec, Copyright (c) 2003 OneRiver Media, Image provided courtesy of Marco Solorio

Figure 1 is a test image that OneRiver Media uses to test a variety of codecs' performance. Figure 2 is the result of successively reencoding the test image ten times using a particular 4:2:2 'uncompressed' codec. As this example demonstrates, even codecs that advertise themselves as uncompressed cause a video image to gradually degrade after successive applications. Blurring, smearing and loss of color information result from the drift caused by eliminating color information in chroma subsampling and its imperfect reconstitution through interpolation by the codec.

In normal video production, a video stream would not be subject to ten successive encoding runs, and artifacting of the severity seen above would not occur. If we are digitizing video with an eye towards its long-term preservation however, we must consider the possibility that digital video streams in our care may be migrated far more often than ten times if we are to keep them accessible to our users. The artifacting seen above is produced by a codec where the software engineers went to a great deal of effort to insure that the compression and decompression operations were, to the extent possible, inverse operations. We can expect more extensive degradation if we subject a video stream to successively *different* compression and decompression operations with different assumptions regarding treatment of the video stream's colorspace.

4:4:4 and/or Bust

It is possible to produce digital video today that is not subject to the degradation demonstrated above by 4:2:2 video. By using 4:4:4 sampling (that is, by eliminating the use of chroma subsampling), it is possible to produce an image that can migrated repeatedly without artifacting. The final example from OneRiver Media below shows the results of ten successive applications of Apple's QuickTime 4:4:4 "None" codec to the test image in Figure 1. This example is a bit-for-bit identical copy of the original test image.

Figure 3-- 10th Generation Copy using Apple 4:4:4 "None" Codec, Copyright (c) 2003 OneRiver Media, Image provided courtesy of Marco Solorio

Using a 4:4:4 sampling regime eliminates one of the problems faced in trying to produce a digital video stream capable of being preserved in the long term, however it raises additional problems. The first of these is the choice of media on which to store the stream. As noted, standard digital video production equipment and digital videotape formats all employ some form of chroma subsampling; even the highest quality equipment and tape formats use 4:2:2 sampling. If you wish to store and retain 4:4:4 video, then, you cannot use videotape systems. You must store the video stream as a file in data repository system of some kind, whether on magnetic disk or using a hierarchical storage management (HSM) system to take advantage of the lower cost of

magnetic tape as a storage medium.

Given the tremendous storage requirements for uncompressed 4:4:4 video, most institutions choosing a storage architecture will probably lean towards an HSM system due to the significant cost savings over an entirely disk-based architecture. For long-term preservation, however, there are at least a few considerations that libraries and archives should bear in mind in deciding upon a storage architecture. First, files on any storage media can become corrupt, and in any video archive of significant size, you will need to automate the process of examining video assets to determine whether they have become corrupted and need to be restored from a backup copy. Software for this purpose is readily available, but the constant checking of all the files in your repository required by this sort of application will place an additional strain on the tape robotics used in most HSM systems, and may hamper performance for your users. Second, any storage architecture you might choose to implement will eventually become antiquated and need to be replaced, and all of your assets migrated to a new system. At that point, the ability to quickly move terabytes or petabytes of information from one system to another will be a paramount concern. While there are tape storage library systems that can sustain the necessary throughput to accomplish this type of migration in a timely fashion, they tend to be on the higher end of the price range for these systems. Those implementing HSM for long-term archival video storage should take care to insure that there is a technologically feasible escape path from whatever storage system they choose to put in place (actually, that caution could equally be applied to those designing disk-based storage systems for video). A final consideration for any storage mechanism is the maximum allowable file size under the operating system used. An hour of standard NTSC 4:4:4 video stored uncompressed will consume 140 GB; for HDTV, an hour will consume over 840 GB. While modern file systems can accommodate files of that size, care must be taken in selecting and configuring a storage system to insure that it can handle the size of files to be generated.

Choosing a file format for storing a 4:4:4 digital stream presents another challenge. While there are several file formats that can store 4:4:4 video, real world production depends on not only the existence of a suitable format, but software capable of supporting it. At the moment, finding a file format that supports 4:4:4 uncompressed video, is publicly documented (preferably in a formal standard), and for which software support is readily available is somewhat difficult. The QuickTime format fulfills the basic technical requirement of supporting 4:4:4 uncompressed video and software support is readily available. While it is a publicly documented format (Apple, 2001), it is also a proprietary one. It provides what is probably the easiest to use format capable of supporting 4:4:4 video today, but archivists employing it may wish to track the availability of public documentation for the format and consider abandoning it if documentation ceases to be available. Motion JPEG 2000 is another alternative, supporting 4:4:4 uncompressed video and also being documented in an official standard (ISO/IEC, 2003). However, software support for Motion JPEG is not as widely available as support for QuickTime. Another possibility that has recently emerged is the Material Exchange Format (MXF), developed by the Pro-MPEG Forum and aimed at the interchange of audio/visual information, along with associated data and metadata, between systems (Pro-MPEG Forum, 2003). It is an XML-based format, and is agnostic

with regards to the video coding and compression algorithms used to store video 'essence'. It has also been submitted to SMPTE for standardization. It has support from several of the major corporations involved in the video production industry, and some early test software is available for working with the format, but the SMPTE version of MXF has not been finalized, and anyone adopting MXF at the moment must be ready to deal with some changes to the official file format when it is standardized.

The Price of Perfection

It is theoretically conceivable to establish a video archive today using 4:4:4 video in a data repository using QuickTime or MJPEG2000 as a file format. However, to do so presents cost issues that will probably be insurmountable for most archives.

As an example, consider the case of New York University Libraries. We have recently completed a survey of our collections to determine the amount of moving image material we possess. As it happens, the amount is considerably more than anticipated (in excess of 30,000 hours), due to the amount of moving image materials in our special collections for which there are no MARC records in our catalog. Let us assume that 10% of that material is at risk, and we wish to preserve it by digitizing it as 4:4:4 video and storing it on a disk-based file repository system, and that we wish to accomplish the digitization in the course of one year.

To accomplish this within those time constraints would require the equivalent of 9 full-time employees for one year working on digitization. As it happens, NYU has a variety of graduate programs with students very conversant with video and film technologies, so our labor costs could probably be reduced by employing student labor. However, labor costs would still approach $350,000 for the year. Equipment costs for digitization workstations capable of the necessary 4:4:4 capture (and quality control on the captured files) would total about $690,000. Captured as 4:4:4 uncompressed video, 3,000 hours of video would consume approximately 425 terabytes of disk storage. Currently, adding a terabyte of storage to our repository system costs approximately $10,000 (the price is somewhat high due to RAID architecture and redundant Fiber Channel paths to the disks), so the total cost of disk storage would be approximately $4,250,000.00. In total, we could anticipate approximate costs to digitize and store 3,000 hours of video of $5,290,000.00.

Unfortunately, that amount is nearly half of our library's current collections budget. We simply cannot afford those costs. While we might be able to rescue a limited amount of material in immediate danger of loss, any large-scale programmatic effort to use lossless digitization of video assets as a preservation strategy is beyond our means. For the moment, we are pursuing a strategy of digitizing items as 4:2:2 video and storing them on Digital Betacam tape, hoping that storage costs will continue to drop and allow us to eventually migrate this material to disk as 4:4:4 video.

"There will be a brief delay"

A crucial piece of information for the strategy NYU is adopting is when we might

reasonably expect to abandon the use of Digital Betacam in favor of storing 4:4:4 video to disk. Essentially, this is a question of when disk storage prices will fall sufficiently to make them competitive with videotape as a storage medium. If it is to be a long wait, excessive production of new Digital Betacam tapes will present us with a significant migration project when we can afford to move video on to disk, and we might wish to consider digitizing materials only when necessary. Fortunately, indications are that the wait for cheaper disk storage will be a short one.

Currently, the media costs for producing two master copies (an original and a backup) for an hour of video on Digital Betacam tape is $70. Adding new disk storage to our repository system costs approximately $10,000.00, so the 143 GB required for an hour of 4:4:4 and accompanying video would cost $1,430.00. If we wish to replicate the video stream on disk for a backup copy, our costs would be $2,860.00.

Between 1992 and 2000, disk storage prices declined at an average rate of 45% per year (Gilheany, 2001). If prices could this rate of decline, by the year 2010, the storage that currently costs $2,860.00 will cost about $79.00, or $9.00 more than the equivalent in videotape today. In short, in a little over five years, it will cost about the same to store 4:4:4 video on disk as it costs us now to store 4:2:2 video on videotape.

Five years is not a particularly long wait, but it does confront libraries and archives with the question of which strategy to pursue at the moment with respect to digitizing video materials. Large scale digitization of video at a level of quality suitable for preservation will be beyond the means of any but the most well-funded of research institutions and a few large commercial entities. Smaller institutions will probably be forced to employ a mix of strategies based upon their available finances, including reformatting some materials to analog media, some to lossy digital, and perhaps a limited amount of extremely valuable and at-risk material to lossless digital formats. NYU is still developing a comprehensive plan for preservation of moving image materials, and our digitization efforts for moving image materials are focused on improving access at this point. However, our choice of Digital Betacam as our initial digital capture format was driven by a desire to choose the 'least lossy' option we can presently afford, anticipating that in a few years we will be able to migrate material digitized now to a truly lossless format.

Bibliography

Apple Computer, Inc. (2001). *QuickTime file format*. Cupertino, CA: Apple Computer, Inc. Retrieved June 19, 2004 from the Apple Computer, Inc. website at http://developer.apple.com/documentation/QuickTime/QTFF/qtff.pdf

The Bancroft Library, University of California at Berkeley (2003). *The Center for the Tebtunis Papyri*. Retrieved June 19, 2004 from http://tebtunis.berkeley.edu/

Gilheany, Steve (2001). *Projecting the cost of magnetic disk storage over the next 10 years*. Retrieved June 19, 2004 from

http://www.archivebuilders.com/whitepapers/22011p.pdf

International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) (2003). *Motion JPEG 2000 derived from ISO base media file format (ISO/IEC 15444-3:2002/Amd 2:2003).* Geneva, Switzerland: ISO/IEC.

OneRiver Media (2003). *OneRiver Media codec resource site.* Retrieved June 19, 2004 from http://codecs.onerivermedia.com

Poynton, Charles (1997). *Frequently asked questions about color.* Retrieved June 19, 2004 from http://www.poynton.com/PDFs/ColorFAQ.pdf

Poynton, Charles (2002). *Chroma subsampling notation.* Retrieved June 19, 2004 from http://www.poynton.com/PDFs/Chroma_subsampling_notation.pdf

Poynton, Charles (2003). *Digital video and HDTV algorithms and interfaces.* New York: Morgan Kaufman Publishers.

Pro-MPEG Forum (2003). *MXF.* Retrieved June 19, 2004 from http://www.pro-mpeg.org/publicdocs/fileinter.html