Transcoding Analysis of a Digital Video Archive

Itapajé Takeguma and Bruno Wanderley

***Abstract*— Digital video presents a great challenge for any entity responsible for its archival: huge amount of data to be stored. This paper presents an empirical research that was conducted aiming to evaluate the introduction of the transcoding technique and the H.264/AVC encoding to the video archival system maintained at Brazilian Senate. Image quality issues were considered, besides storage reduction, but there was no study of audio degradation or CPU cost.**

Keywords—Transcoding, video archive, H.264/AVC

I. INTRODUCTION

Senado Federal do Brasil (Brazilian Senate) captures and broadcasts video from its plenary, commissions and other events. The responsibility for archiving this material is held by Senate's *Centro de Documentação Multimídia* (Multimedia Document Center), which currently stores about 45.000 hours of video, including gatherings since 1992.

The DV25 codec is the current choice, not only for *TV Senado*'s internal workflow, including playout, but also for the storage inside the archival system. Considering the 28.8 Mbit/s throughput of this codec, the total amount required for storing that video asset is about 580 TB, besides backup.

It is true that DV25 presents some interesting features that place it as good video codec candidate for a TV workflow, for instance: constant bitrate and intra frame only compression. While the former guarantees a predictable storage and network usage, the latter facilitates video positioning, clipping and (reverse) playback.

On the other hand, its 28.8 Mbit/s bitrate (25 Mbit/s for video stream) seems to represent a huge amount of data for archiving Standard Definition (SD) video. Using similar approach, one might select DV100 to store 1080p video, which would increase four times storage and network load.

Since most of these archived events are shot by fixed cameras, which occasionally pan, tilt and zoom, focusing subjects who speech in from of still background, it is reasonable trying some inter frame compression-capable codec.

The objective of this research was to evaluate the usage of H.264/AVC into *Senado*'s archival system. Both the DV25 archived and raw captured videos were (re) encoded to the H.264/AVC standard and then returned back to the DV25. The resulting video of this last transcoding was compared against the archived DV25 (both objectively and subjectively) in order detect quality loss.

The following figure presents the transcoders elements that would be required to be introduced to *Senado*'s archival system.

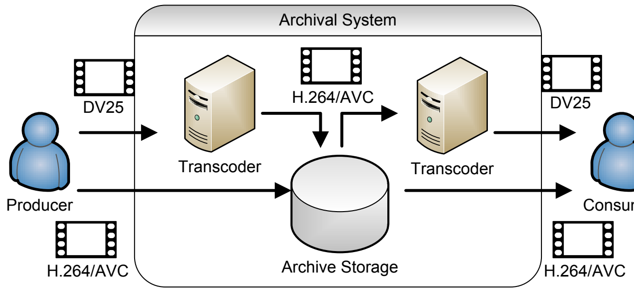

Fig. 1. Proposed archival workflow structure.

The reasoning behind the first transcoding, i.e. moving to H.264/AVC, is to save storage resources, but could also be used to improve quality, once H.264/AVC introduces far more compression techniques than DV25.

The latter transcoding, which converts the video back to DV25, is a legacy requirement: given that the document is converted to this standard, *TV Senado*'s internal processes can continue without even notice the transcoding step.

Also, it is reasonable to forecast a shortcut: archiving or retrieving H.264/AVC encoded video directly (lower arrows at Fig. 1). Clearly, avoiding unnecessary transcoding steps can save processing time and also improve video quality.

II. BACKGROUND

What video codec to use at an archival system can generate an endless debate, some might recommend using DV25 in the broadcast industry. Ljubomir Jovanov et. al. [5] has compared SMPTE D10 and DV25. In their experiments, the SMPTE D10 demonstrated superior image, though it used double the bitrate. In order to objectively evaluate the image quality, they calculated the peak signal-to-noise ratio (PSNR) and the Structural Similarity (SSIM).

Wang [12] advocates that SSIM works closer to the human vision than other indexes. In his work, he compares the SSIM to the mean squared error (MSE) and presents some images that have similar MSE, but have very different visual quality.

Pearson [7] has evaluated Motion JPEG 2000 (MJ2) and its lossless compression. He reminds how video archivists are interested in the subject of lossless video coding. In his work, he was able to reach a 3.33:1 compression, a little worse than DV25 that is 5:1.

Wootton [15] recommends that archival systems use uncompressed source video, if affordable. He believes that when the video reaches the consumer, it has been compressed by a factor of 10 to 1. But it would also be acceptable to use a "*nearly lossless storage format*".

Itapajé de Farias Seto Takeguma, Secretaria Técnica de Eletrônica, Senado Federal do Brasil, Distrito Federal-DF, Brazil, E-mail: itapaje@senado.gov.br.
Bruno Wanderley, Universidade Federal Fluminense, Niterói-RJ, E-mail: bruno@unisat.com.br.

Whichever codec is chosen for encoding the archived videos, be it lossy or lossless, transcoding could be applied to a video archival, aiming to easy the consumer access or, as presented here, to save storage resources.

In the context of the Open Archival Information System (OAIS) Reference Model [8], when an archive modifies the internal document data, it is applying a migration known as transformation. It could be reversible (e.g. if applied a zip compression), or irreversible.

There is an associated risk of losing information while transforming data, and that risk must be assessed and might be accepted, if the long term preservation of the document is in danger. It must be done in a way that "*there is no significant information loss to the Designated Community*" [8].

That reference model also describes three different arrangements for document packages: Submission Information Package (SIP), which represents data and metadata, prepared by the producer; Archival Information Package (AIP) that is how the archive maintains the information; and finally the Dissemination Information Package (DIP), which is prepared by archive and delivered to the consumer.

In that sense, the transcoding can be used to make sure that *TV Senado* can submit (SIP) e retrieve (DIP) video in DV25, but the archive can store (AIP) H.264/AVC encoded media.

Ahmad et. al. [4] present the transcoding operations and their classifications: in the context of that work, the transcoding proposed in this paper would be classified as heterogeneous and between standards.

They also show typical transcoder architectures, but as the FFmpeg software [3] did much of that work transparently, this subject was not studied here. As DV25 does intra frame compression only, FFmpeg should work as a cascaded transcoder, as no motion compensation information would be available to the encoder.

III. RESEARCH

The experiments conducted were based on samples of the *Senado*'s Plenary of July 20^{th} 2012, including the DV25 archived version and a parallel raw capture. Samples of *TV Senado*'s *Inclusão* program were also included, but only the DV25 version. This production contained images from Copacabana beach and could present harder challenges, compared to the Plenary.

These video sources were clipped to smaller samples of 30 seconds, which were arranged into 6 groups, named A to F, that contained 2 video clips each, as show in the Table I.

TABLE I. *VIDEO CLIPPING GROUP ORGANIZATION.*

Group	**Content**	**Clip 1**	**Clip 2**
A	Plenary	Archived DV25	DV25 transcoded from H.264/AVC at 1.5 Mbps, originated from raw video. Henceforth called **raw>AVC-1.5>DV25**.
B	Plenary	Archived DV25	DV25 transcoded from H.264/AVC at 15.0 Mbps, originated from raw video. Henceforth called **raw>AVC-15.0>DV25**.
C	Plenary	Archived DV25	DV25 transcoded from H.264/AVC at 1.5 Mbps, originated from archived DV25. Henceforth called **A-DV25>AVC-1.5>DV25**.
D	Plenary	Archived DV25	DV25 transcoded from H.264/AVC at 15.0 Mbps, originated from archived DV25. Henceforth called **A-DV25>AVC-15.0>DV25**.
E	*Inclusão* Program	Archived DV25	DV25 transcoded from H.264/AVC at 1.5 Mbps, originated from archived DV25. Henceforth called **A-DV25>AVC-1.5>DV25**.
F	*Inclusão* Program	Archived DV25	DV25 transcoded from H.264/AVC at 15.0 Mbps, originated from archived DV25. Henceforth called **A-DV25>AVC-15.0>DV25**.

There was always an archived DV25 clip in each of these groups, but groups A and B included DV25 transcoded from H.264/AVC originated from raw video, while in groups C to F, the H.264/AVC originated from the archived DV25 itself. Also, the bitrate was changed from group to group.

The lower bitrate (1.5 Mbps) would be recommended by some authors [2] for SD video presenting medium motion. The higher bitrate (15.0 Mbps) was an extrapolation of the lower one.

As explaining word by word the transcoding path may last lots of words and can make the text harder to read, it will be given synonyms for the video, according to the steps that were followed to obtain it. In that sense, the "raw>AVC-1.5>DV25" means that it all started with a raw capture, which was encoded to a 1.5 Mbit/s H.264/AVC and finally transcoded to DV25.

In order to objectively evaluate the transcoded video image quality, the Structural Similarity (SSIM) index was chosen and implemented in Java™ [4]. This index was also useful to discover encoding parameters that would yield a reasonably high SSIM value, at an acceptable video size.

Many of the steps followed in this research depended on the open source software FFmpeg: capturing raw video, clipping, transcoding and dumping video frames. These frames where presented to that Java™ tool to compute the SSIM.

For the subjective evaluation, the video groups were organized in files and named A1.dif, A2.dif, B1.dif… F2.dif. A Group A video, for instance, would occupy position A1.dif or A2.dif randomly. Both the video files and a questionnaire were distributed to 19 evaluators. They were asked to point the best clip for him or her, for each group.

IV. RESULTS

The answers to the questionnaires given by the 19 evaluators were consolidated in the following histogram.

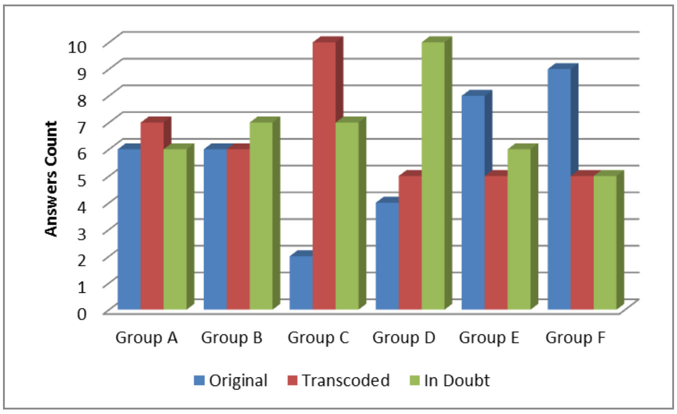

Fig. 2. Distribution of preferred video, as answered by evaluators.

Groups A and B, whose transcoded video originated from raw video, presented the most equivalent results. The rest of the groups presented somewhat scattered values, but in groups E and F most of the evaluators preferred the archived DV25 video.

Because the evaluator could get confused, he or she was asked to decide between only two videos. The objective evaluation could be executed on every video clip. For groups A to D, the raw video could also be included in the analysis, as displayed in the following graph.

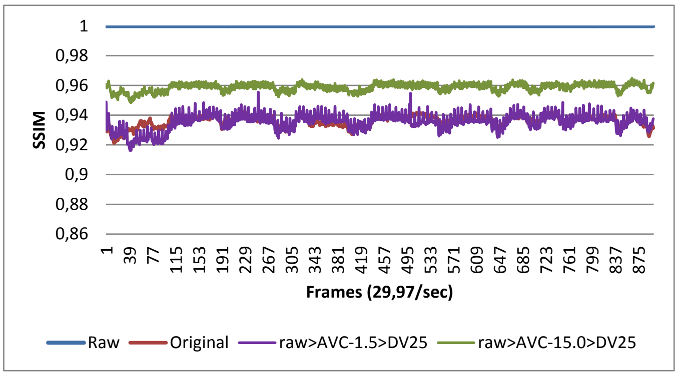

Fig. 3. SSIM values for group A, with the original raw video as reference.

The raw video is the reference one, and so it's SSIM value are allways 1 (one). Just bellow it comes the raw>AVC-15.0>DV25, presenting a mean value of 0.9558 and stadard deviation of 0.0027. And then, almost in a tie, the archived DV25 and the raw>AVC-1.5>DV25; the mean value is respectively 0.9361 and 0.9351, while the standard deviation 0.0037 and 0.0057.

The higher standard deviation presented by the raw>AVC-1.5>DV25 confirms the comb shape of is grapth. The 15.0 Mbit/s version presented a lower standar deviation when compared to the archived DV25 and presents a steadier graph.

One might guess, from the previos graph, that the H.264/AVC would generate a better result than DV25, for archiving purposes. It must be observed that the raw video and the archived DV25 came from different machines, that would probabily generate sligtley unequal results. The following graph presents that same group, but, at this time, the reference video is the archived DV25.

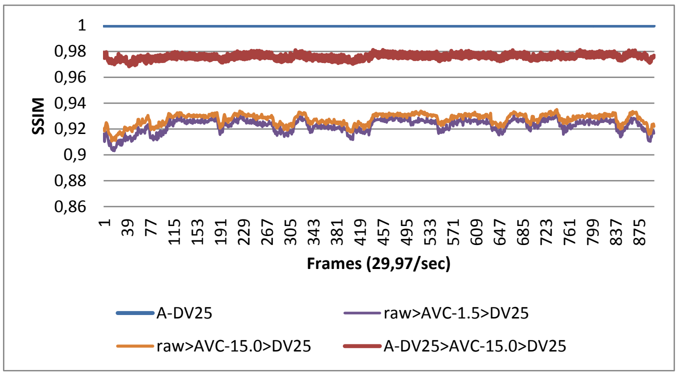

Fig. 4. SSIM values for group A, with the original DV25 as reference.

At this graph, the transcoded videos originated from raw capture present lower scores then they did in Fig. 3. Though not presented to evaluators, an A-DV25>AVC-15.0>DV25 was forged aiming to enrich the comparison.

That video presented better results than the other transcoded videos (originated from raw source). Once Raw and A-DV25 are themselves different videos, it is expected that each corresponding derivation be more similar to its transcoding source.

From these previous graphs and the subjective evaluation, it seems that some degree of dissimilarity presented by the SSIM, does not necessarily imposes noticeable quality loss. Wang [13] has presented slightly different images, which were modified changing contrast or luminance, and shows small SSIM change.

As group B graphs are similar those previously shown, they were omitted. Group C and D graphs present a more visible comb effect, but the latter also displays a stepped graph, as presented in the following figure, so group C graph was also omitted.

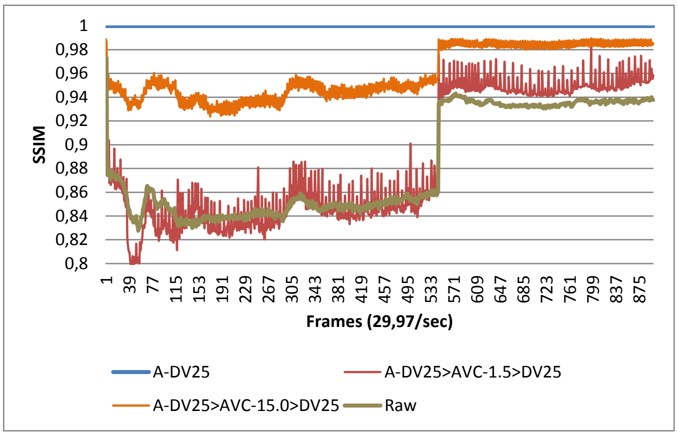

Fig. 5. SSIM values for group D.

The SSIM shows a significant improvement at frame number 547. Before that frame, the mean value for the A-DV25>AVC-15.0>DV25 was 0.9427 and then it rose to 0.9843, while the A-DV25>AVC-1.5>DV25 rose from 0.8445 to 0.9498.

Exactly between frames 546 and 547 there is a camera change, but the exact nature of the step at the graph could not be uncovered: it might be due signal noise or different challenges presented by each scene.

Despite that step, since the raw video SSIM (also plotted in Fig. 5) presents dissimilarity close to the video transcoded from the AVC-1.5 and, considering that in group D questionnaire, the "in doubt" answer prevailed, that step might not be noticeable to the evaluator.

The F group (edited video), as presented below, displays high deviation from the reference video at certain points. To avoid a graph flattening, different scales had to be used.

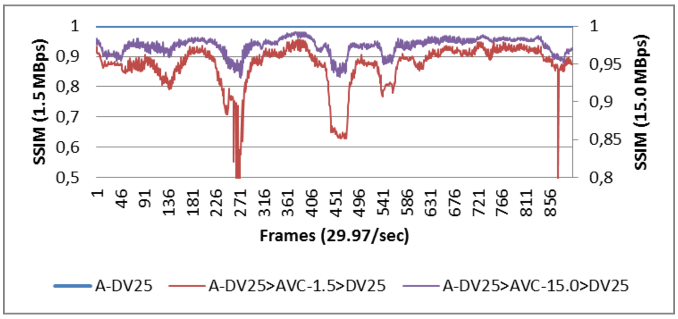

Fig. 6. SSIM values for group F.

Whenever the SSIM stepped down, it was precisely when it was in course a dissolve transition. This video effect consists of gradually transitioning from one scene to another. Wottoon [15] discusses about this subject:

A consequence of this kind of scene transition [dissolve] *is that every pixel changes value on every frame, and any motion compensation is also compromised since subjects are partially transparent and may be moving in completely different directions in the two scenes.*

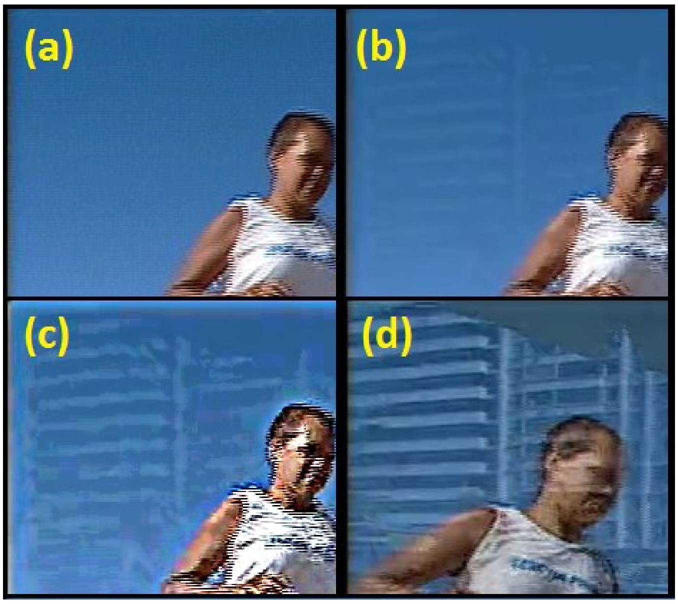

Fig. 7. Group F clippings: (a) 272 frame from archived DV25 video and (b) A-DV25>AVC-1.5>DV25; (c) same as b, but applying unsharp mask; (d) 285 frame from archived video.

The previous figure was assembled using frames from group F at frame 272 (Fig. 7 a, b and c), when a dissolve effect was up to start, and then frame 285 (Fig. 7 d), that was just in the middle of that effect. Sub pictures b and c came from the same transcoded DV25, but to the latter was applied an unsharp mask.

While the building that is being dissolved (visible in frame 285) does not appear in the archived DV25, the transcoded video anticipates that object which is some frames ahead. This issue can easily disturb the SSIM computation.

When the encoder is programmed for inter-frame compression, the dissolve effect presents a real challenge as explained above. Wottoon suggests that using a different transition, such as wipe, would result in better compressions, though, for video editors, wipe and dissolve have different meanings.

The minimum, maximum, mean and standard deviation values calculated for the SSIM were consolidated and presented in the following table.

TABLE II. *MINIMUM, MAXIMUM, MEAN AND STANDARD DEVIATION SSIM VALUES CALCULATED BETWEEN TRANSCODED AND ARCHIVED VIDEOS.*

	DV25 Transcoded from H.264/AVC at 1.5 Mbit/s				DV25 Transcoded from H.264/AVC at 15.0 Mbit/s			
Groups	Min. Value	Max. Value	Mean	Std. Deviation	Min. Value	Max. Value	Mean Value	Std. Deviation
A	0,903	0,932	0,923	0,005	0,911	0,935	0,927	0,004
B	0,904	0,932	0,920	0,007	0,911	0,935	0,925	0,006
C	0,916	0,977	0,941	0,009	0,972	0,987	0,980	0,003
D	0,793	0,983	0,886	0,054	0,924	0,990	0,959	0,022
E	0,279	0,970	0,896	0,072	0,692	0,987	0,970	0,037
F	0,424	0,960	0,868	0,074	0,932	0,992	0,973	0,011

The table shown above confirms that groups A, B, C and D present superior results when compared to the edited videos from groups E and F. Also, the higher standard deviation presented when the video was transcoded from H.264/AVC at 1.5 Mbit/s justify that combing effect presented at those graphs.

As discussed previously, group D presented a stepped graph, so its standard deviation is significantly higher than group C, which was assembled similarly.

The following graph presents the final results that could be reached by introducing the transcode technique to that video archival maintained by Brazilian Senate.

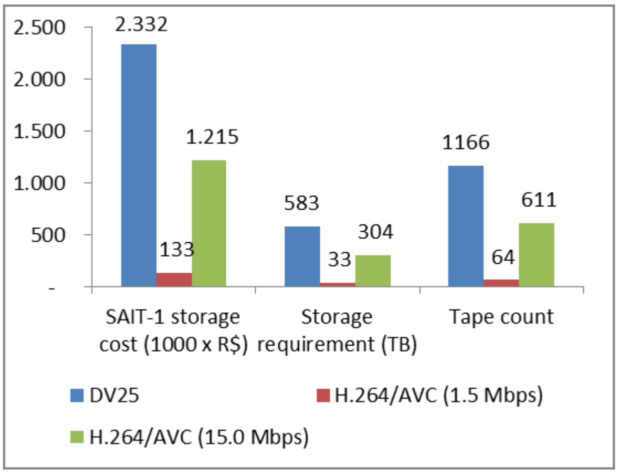

Fig. 8. Estimated resource saving by introducing transcoding.

A high financial saving would be expected, as the current storage media is the discontinued SAIT-1. Once this technology is swapped, the gigabyte cost will fall, and so will the economic impact of the transcoding. Though, the storage requirements and the tape count would maintain the relative drop.

An overall 48% of cost cut would be expected, considering applying transcode to every archived video, including those edited programs. As the majority of documents is composed of legislative events (which use no transition effect), the estimated resource savings consider transcoding every archived video.

V. CONCLUSIONS

Video archiving requires huge amount of data storage, and as display technology evolves from Standard Definition to High Definition and then to Ultra High Definition, also will increase storage demand. Fortunately, as physical media technology evolves, the gigabyte price has steadily fallen.

Although paying for technology upgrade may in fact save money after some months, for Brazilian Public Sector, moving to the next generation may last long enough for the upgrade to become outdated.

The results presented in this research show that the archived DV25 and the transcoded one have almost indistinguishable visual quality, unless for edited video. While transcoding legacy videos to the H.264/AVC can save at least 48% of storage resources, encoding directly to H.264/AVC would generate similar economy, but leave an even more unnoticeable trail.

The transcoding technique can also be applied directly, without hardware expenditure, though hardware investment would be very welcome, as these video operations present intensive processing requirements.

Finally, the combination of subjective and objective analysis creates a solider confidence about H.264/AVC archiving than the mere faith on DV technology that some might have.

ACKNOWLEDGEMENTS

This paper is derived from a MBA program on Digital TV, supported by Brazilian Senate and lectured by Universidade Federal Fluminense (UFF) professors.

Authors would like to thank those 19 evaluators, who spent their time on the subjective assessment: Furtado, M.; Azevedo, P.; Cavalcanti, D.; Dobke., M.; Ginez, R.; Ferreira, L.; Santos Júnior, J..; Ramos, K.; Bicalho, D.; Oliveira, D.; Cavalhero, M.; Amorim, R.; Caldas, E.; Montalvão, L.; Manhães, E.; Freitas, F.; Costa Santos, C.; Guimarães, F.; Toscano, R.

FUTURE WORK PROPOSITION

This research involved DV25 encoded video at standard definition that is still currently produced by *TV Senado*. Similar work can be conducted involving high definition (HD) video and H.264/AVC or the recently launched H.265/HEVC.

Besides, audio degradation, while not formally studied, was noticeable indeed after some transcoding iterations. Some effort could be spent on assessing audio codecs, lossy and lossless.

REFERENCES

[1] I. Ahmad, Video Transcoding: An Overview of Various Techniques and Research Issues. IEEE transaction on multimedia, vol. 7, no. 5, october 2005.

[2] K. Amerasinghe, H.264 for the rest of us. [Online] Available: http://wwwimages.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/video/articles/h264_primer/h264_primer.pdf

[3] FFmpeg. [Online] Available at http://ffmpeg.org.

[4] Java. [Online] Available: http://java.com.

[5] L. Jovanov, et. al. Quality assessment of D10 and DV25 video codecs for broadcasting purposes.

[6] J. Jung, 19 ffmpeg commands for all needs. Blog containing useful information for developers. [Online] Available: http://www.catswhocode.com/blog/19-ffmpeg-commands-for-all-needs.

[7] G. Pearson and M. Gill, An Evaluation of Motion JPEG 2000 for Video Archiving, Proc. Archiving 2005 (April 26-29, Washington, D.C.), IS & T (www.imaging.org), pp. 237-243.

[8] Reference Model For an Open Archival Information System (OAIS), june 2012. [Online] Available: http://public.ccsds.org/publications/archive/650x0m2.pdf.

[9] G. Sullivan, et al. H.264 The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions. August 2004.

[10] H. Sun. Digital Transcoding for Transmission and Storage. Boca Raton, Florida: CRC Press, 2005.

[11] Z. Wang; A. Bovik. A Universal Image Quality Index. IEEE Signal Processing Letters, Vol 9, pp. 81-84, March 2002.

[12] Z. Wang, et al. Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing, Vol. 13, No. 4, April 2004.

[13] Z. Wang; A. Bovik, Mean Squared Error: Love It or Leave It – A new look at signal fidelity measures. IEEE Signal Processing Magazine, Vol. 26, No. 1, pp. 98-117, January 2009.

[14] T. Wiegand, et al. Overview of the H.264/AVC Video Coding Standard. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 7, July 2003.

[15] C. Wootton. A Practical Guide to Video and Audio Compression. Burlington: Focal Press, 2005.

[16] Xin, J., et al. Digital Video Transcoding. Proceedings of the IEEE, Vol. 93, No. 1, January 2005.