INTRODUCTION TO JPEG XS – THE NEW LOW COMPLEXITY CODEC STANDARD FOR PROFESSIONAL VIDEO PRODUCTION

Joachim Keinert[1], Jean-Baptiste Lorent[2], Antonin Descampe[2], Gaël Rouvroy[2], Siegfried Fößel[1]

[1] Fraunhofer IIS, Germany and [2] intoPIX SA., Belgium

ABSTRACT

Due to increasing resolutions and 360° capture, broadcast and video production is characterized by handling large data volumes. To ease such data intensive workflows a novel image and video codec called JPEG XS is currently standardized. It focuses on film production, broadcast and Pro-AV markets and excels by ultra-low latency and ultra-low complexity. Moreover, it permits multiple encoding and decoding cycles with minimum quality loss and it can be implemented on different platforms such as CPU, GPU, FPGA and ASIC.

The paper describes these properties in more detail and explains how they help to devise optimal video and transmission workflows. Moreover, it details the used technologies in form of a block diagram and summarizes the results of the quality evaluation ensuring visually lossless compression.

INTRODUCTION

The goal of better visual experiences in the form of higher resolution and 360° movies causes transmission throughput in production networks to increase at a larger pace than the available network infrastructure. This holds both for legacy infrastructures, whose replacement by a new generation is very costly, as well as for IP networks needing to simultaneously route multiple video streams.

While standard video compression could be thought to solve these challenges, in practice existing standards such as JPEG, JPEG 2000 or HEVC do not comply with the needs of the film and broadcast production networks, as they are not designed for ultra-low latency and low complexity while achieving visually lossless compression. Consequently their implementation costs are too high to justify their application. To overcome this situation, the JPEG Committee (formally known as ISO/IEC SC29 WG1) has been starting the standardization of a novel compression codec called JPEG XS.

USE CASES AND TARGET MARKETS

While JPEG XS is generally usable in all applications requiring low complexity and low latency compression, it has been specifically designed to meet the requirements of live productions, broadcast and digital cinema workflows, Pro-AV markets, keyboard-videomouse (KVM) extender applications, as well as Virtual Reality (VR) gaming. The following subsections will discuss the related use cases and explain their specific needs to illustrate the core application scenarios foreseen for the JPEG XS codec.

Live Video Transmission (Streaming)

In order to reduce the costs of live video transmissions, JPEG XS is intended to be applicable for all scenarios where today uncompressed images are transmitted over either existing legacy infrastructures, or future IP production networks. To this end, visually lossless compression quality is as important as robustness to multiple encoding and decoding cycles, such that several devices each compressing and decompressing the signal can be chained. Massimo and Hoffmann (1) recommend that codecs should maintain the image quality for at least seven compression-decompression cycles. Moreover, as pointed out by Cronk and Meyer (2), the additional latency introduced by one coding and decoding cycle should be below a couple of lines in order to avoid any human-perceptible delay between signals processed by different processing chains.

The target compression rates to be supported by the codec can be derived from the typical image sizes, frame rates and available infrastructure as explained in Table 1.

Table 1: Target compression ratios

video stream	video throughput	target physical link	available throughput[1]	compr. ratio
2K / 60p / 422 / 10 bits	2.7 Gbit/s	HD-SDI	1.33 Gbit/s	~ 2
2K / 120p / 422 / 10 bits	5.4 Gbit/s	HD-SDI	1.33 Gbit/s	~ 4
4K / 60p / 422 / 10 bits	10.8 Gbit/s	3G-SDI	2.65 Gbit/s	~ 4
2K / 60p / 422 / 10 bits	2.7 Gbit/s	1G Ethernet	0.85 Gbit/s	~ 3
2K / 60p / 444 / 12 bits	4.8 Gbit/s	1G Ethernet	0.85 Gbit/s	~ 6
4K / 60p / 444 / 12 bits	19 Gbit/s	10G Ethernet	8.5 Gbit/s	~ 2.2
2x [4K / 60p / 444 / 12 bits]	37.9 Gbit/s	10G Ethernet	8.5 Gbit/s	~ 4.5
8K / 120p / 422 / 10 bits	85 Gbit/s	25G Ethernet	21.25 Gbit/s	~ 4

Compressed File Based Workflows

Similar to live video streaming, also file based workflows can benefit from a low complexity compression by preventing the data transfer within the network to be the bottleneck. Moreover, in case the camera already creates compressed images, the ingest time can be significantly reduced. In order to be effective, fast software encoding and decoding is necessary. Additionally, multiple coding cycles must deliver the same quality than a single compression and decompression operation.

Further Applications

Besides the use cases mentioned above, the JPEG XS codec is also intended to be applied for the transmission of video signals between head mounted displays and the

[1] On Ethernet links, a 15% overhead has been taken into account

image generating source computer, requiring thus very low latencies. Low complexity frame buffer compression in display devices or video encoders finally permits to reduce the system costs, or even reduce the power consumption in case of embedded devices.

TARGET PLATFORMS

To support the above-mentioned use cases, the JPEG XS codec needs to allow real-time implementations on many different platforms, such as FPGAs, CPUs, GPUs and ASICs. This imposes strict constraints on the compression algorithm to be used, since all those platforms have quite different properties.

Single core CPU implementations, for instance, only offer a restricted fine grain parallelism by SIMD instructions. Multicore implementations, on the other hand, suffer from synchronization overhead when the granularity of the parallelism is too small. FPGAs excel by a large amount of fine grained parallelism and thus can outperform CPUs. On the other hand, achievable clock frequencies are limited and only a fraction of an image can be stored at a given time to avoid external memories. Finally, GPUs can be extremely fast, but need a massive amount of parallelism in order to be efficient.

Hence, to optimally support the different target platforms, JPEG XS needs to provide both coarse grained and fine grained parallelism. Even more important, it needs to be possible to encode and decode one and the same bit stream in real-time on FPGAs, CPUs, GPUs and ASICs, although the latter have very different properties.

KEY PROPERTIES OF THE JPEG XS CODEC

Based on the use cases described above, the following target key properties have been defined for the JPEG XS codec:

- Visually lossless for compression ratios up to 6:1 for both natural and screen content and for at least seven encoding and decoding cycles
- Low complexity implementation on CPU, GPU, FPGA and ASIC avoiding any serial bottleneck in the encoding and decoding process
- Latency of one encoding and decoding cycle shall not exceed 32 lines.
- FPGA implementations should not require any external memory and should not occupy more than 50% of Artix7 XC7A200T or 25% of a Cyclon5 5CEA9 when applied to 4k60fps 4:4:4 video content (with 8-bit colour precision).
- An i7 processor should be able to run an optimized software implementation in real-time for 4k 4:4:4 8-bit 60p content

COMPARISON WITH STATE OF THE ART ALGORITHMS

Based on the requirements above, it is easy to see that existing standards do not comply with the needs of film and broadcast production networks. JPEG-LS [6] and JPEG [5] as well as its successor JPEG-XT [7], which provides backward compatible support of higher bit depths, make a precise rate control difficult, and typical implementations show a latency of one frame. JPEG 2000 [8] uses a complex entropy coder, implying many hardware and software resources for real-time implementations. HEVC [9] as a distribution codec needs a huge encoding complexity without ensuring multi-generation robustness. VC-2 [10] on the other hand is of low complexity, but the applied technology only delivers limited image quality. ProRes as documented by a SMPTE disclosure document [11] is based on macro blocks of 16x16 pixels, making a low latency implementation below 32 lines impossible. Moreover, the symbol wise entropy coding makes fast CPU implementations challenging. DSC [12] finally targets ASIC-based display compression, making efficient implementations on FPGAs and GPUs hard to achieve.

Due to the shortcomings of the existing codecs, the JPEG committee has elaborated a novel low complexity codec called JPEG XS, that provides a precise rate control with a latency below 32 lines and that fits in a low cost FPGA. The compression quality was requested to be superior to VC-2 while supporting implementation on different platforms.

The following section will describe the key technology components that differentiate it from existing standards and ensure compliance with the collected requirements.

ARCHITECTURE OF THE JPEG XS CODEC

Block Diagram

Figure 1 shows the overall block diagram of the JPEG XS codec. In case of RGB input, the colour components are decorrelated by means of a lossless colour transform identical to the one used in JPEG 2000. Next, an integer irreversible wavelet transform is applied. In order to comply with the latency constraints and to avoid excessive memory requirements, only up to two vertical wavelet decompositions are envisaged. In horizontal direction, up to five successive decompositions are permitted.

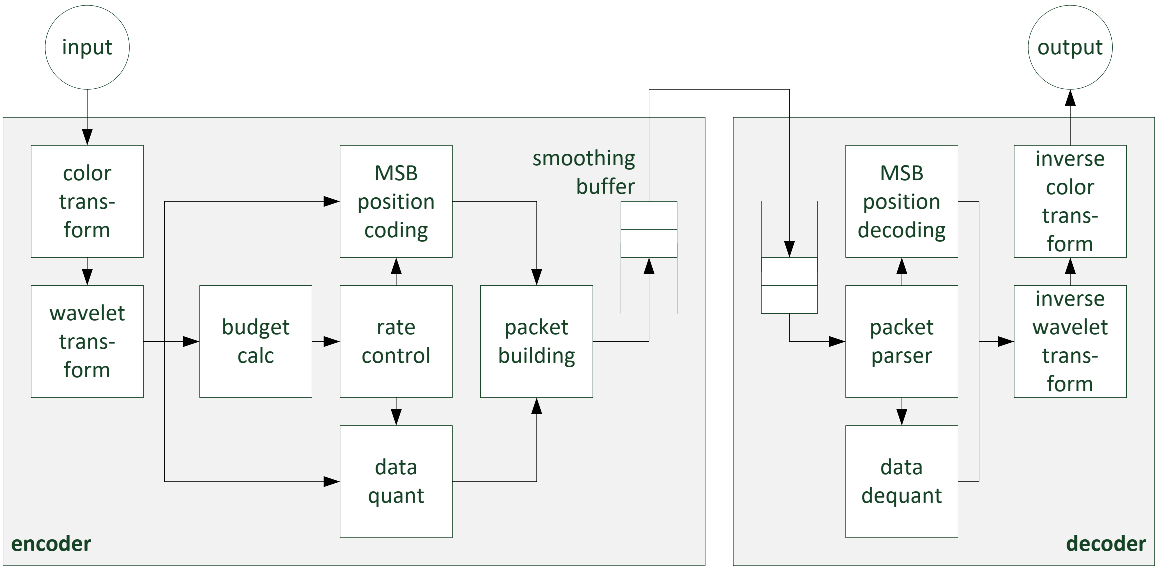
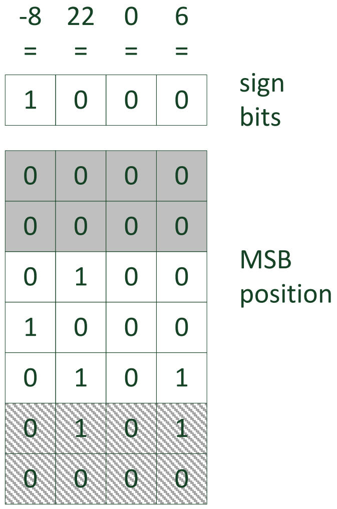


Figure 1: Architecture of the JPEG XS encoder and decoder

The resulting wavelet coefficients are analysed by a budget computation module that predicts the number of bits required for each possible quantization. Since larger quantization means heavier signal distortion, the rate control algorithm computes the smallest quantization factor that does not exceed the bit budget available for coding the wavelet coefficients. Then the wavelet coefficients are entropy coded as described in the next section. Finally, all data sections are combined into a packet structure and sent to the transmission channel. A smoothing buffer ensures a constant bit rate at the output of the encoder, although the input image might consist of input regions that are easier to compress, and others that require more bits per pixel.

Given that the decoder should be able to process the pixels with a constant clock frequency, the number of bits read per time unit varies depending on whether a current wavelet coefficient is easy to compress or not. These rate variations are again compensated by a smoothing buffer at the input of the decoder. A packet parser splits the bit stream into individual data chunks representing parts of a sub-band before the wavelet coefficients are decoded and transformed back into the spatial pixel domain.

Entropy coding



-8	22	0	6
=	=	=	=
1	0	0	0
0	0	0	0
0	0	0	0
0	1	0	0
1	0	0	0
0	1	0	1
0	1	0	1
0	0	0	0

Figure 2: coefficient group for entropy coding

In order to be able to represent an image with as little bits as possible, it is crucial to represent frequently occurring pixel values by short code words, while rare pixel values can be represented by larger code words. This process is called entropy coding.

Unfortunately, coding and decoding such variable length words requires significant hardware and software resources. In order to allow low complexity implementations, it has hence been decided to perform the variable length coding not on coefficient granularity, but on a group of four coefficients. Figure 2 shows such a coefficient group. Each coefficient is represented by a sign bit and a fixed number of magnitude bits. Entropy coding can then be performed by omitting the leading zero bit-lines of each coefficient group (plain grey lines on Figure 2). To this end, the encoder signals for every coefficient group the so called MSB position. It corresponds to the most significant bit-line of the coefficient group, where at least one coefficient bit equals one. To encode this MSB position value, it is first subtracted from the one in the horizontal or vertical neighbouring coefficient group, and this difference is then encoded by a variable length code.

This mechanism permits to only embed bit-lines with lower significance than the leading zero bit-lines into the bit stream, leading to a data reduction. In case the available bit budget is not sufficient to include all bit-lines, their number can be reduced by quantization.

QUALITY EVALUATION

In order to validate that the newly developed codec delivers the image quality required by professional applications, the JPEG committee has elaborated a set of evaluation methodologies specifically devoted to visually lossless compression [3][4]. The following sections summarize the most important results.

Objective Single Cycle Quality Evaluation

In order to cover a large number of different images, an objective evaluation has been performed, computing the *Sequence Peak Signal to Noise Ratio* (SPSNR) between M original images C_k ($k \in [1, M]$) and the images C'_k compressed and decompressed once.

$$SPSNR = 10 \cdot \log I_{max}^2 - 10 \cdot \log \sum_k \sum_c \left(\frac{1}{w_c \cdot N_c(k) \cdot M} \cdot \sum_{(i,j)} \left(C_k(i,j,c) - C'_k(i,j,c) \right)^2 \right).$$

I_{max} is the maximum possible sample value, $N_c(k)$ the number of samples of component c, and w_c the colour weighting factor. It equals $w_{1,2,3} = \frac{1}{3}$ for 4:4:4 content while 4:2:2 content uses $w_1 = \frac{1}{2}$ and $w_{2,3} = \frac{1}{4}$. The test set contained natural and computer generated images as well as screen content as depicted in Table 2.

Table 2: One image per sequence used for objective quality evaluation

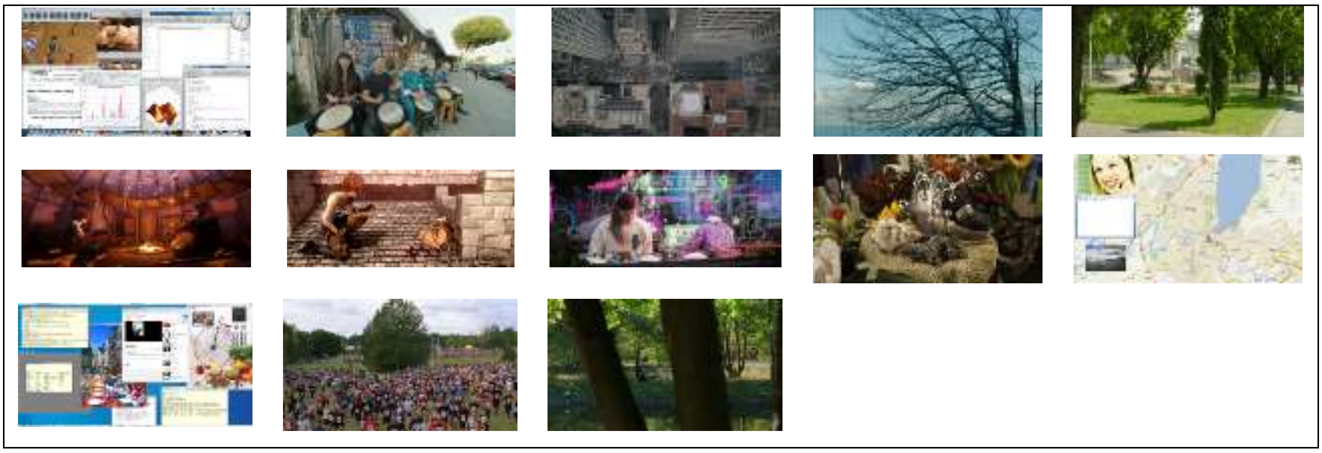


Figure 3 exemplarily depicts the SPSNR obtained for all 4:4:4 8-bit sequences (left) and all 4:2:2 10 bit sequences (right) of Table 2. It demonstrates that JPEG XS in its current development state is almost as good as a tile-based JPEG 2000 codec fulfilling the latency constraints. The latter, however, has much higher complexity. JPEG XS is also better than VC-2, despite the latter is extended by a colour transform not foreseen by the SMPTE standard.

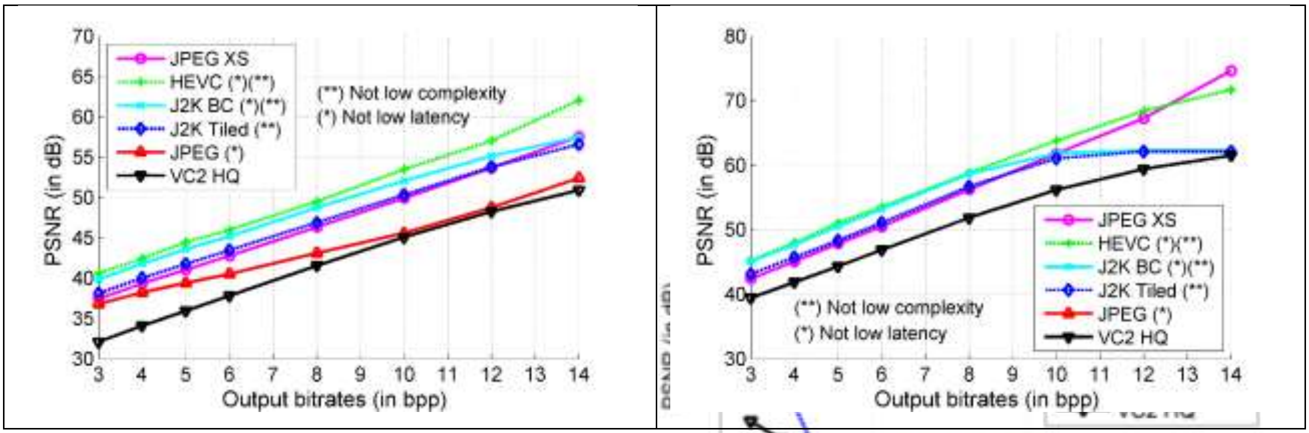


Figure 3: SPSNR for 4:4:4 8-bit (left) and 4:2:2 10-bit images (right). Actual configurations of each codec are described in more details in [13]

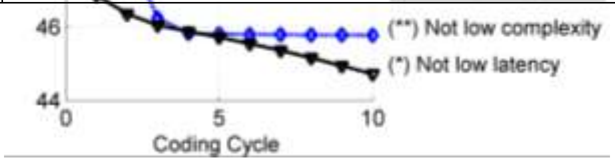


Figure 4: SPSNR for concatenated 4:2:2 10-bit sequences

On the other hand, it can also be seen that applications not needing a low latency or a low complexity implementation but requiring maximum compression are better served by traditional codecs such as HEVC or the JPEG 2000 broadcast profile (J2K BC in Figure 3).

Objective Multi Cycle Evaluation

In order to ensure that multiple encoding and decoding cycles do not accumulate quality losses, the SPSNR has also been computed for 10 encoding and decoding cycles. The results in Figure 4 demonstrate that the JPEG XS codec maintains the initial quality and can thus be used for compressed workflows in broadcast and video production pipelines. All other codecs show much stronger quality decay, causing that JPEG XS delivers the best quality from all codecs starting from the 4th coding cycle.

Subjective Multi Cycle Quality Evaluation

As objective metrics cannot accurately predict the perceived quality, a subjective flicker test with the most challenging images has been performed [3]. To this end, the screen has been split into two regions, showing two times the same image crop. While one side of the split screen showed only the original image, the other side flickered between the original and the compressed one, whereas the observer did not know which side was flickering. Then the observers were asked to identify the flickering side. Querying a large number of observers then allows computing the following quality metric:

$$q = 2 \cdot \left(1 - \frac{n_{ok} + 0.5 \cdot n_{nd}}{n_{total}}\right) \tag{1}$$

The variable n_{ok} represents the number of ratings where the observer correctly identified the flickering side, and n_{nd} the number of ratings where the observer could not decide between the two sides. Since just guessing which side is flickering should statistically lead to $n_{ok} = 0.5 \cdot n_{total}$, a quality value of $q = 0$ means that all observers could detect the compression artefacts, while $q = 1$ corresponds to perfect visual lossless compression. As human beings can detect flickering very well, such a test is very sensitive to any kind of compression artefacts, including colour shifts.

As such quality evaluations are extremely time consuming, those tests have so far only been done at the very beginning of the standardization process. They hence do not reveal the latest codec improvements, but can only indicate a trend.

Figure 5 depicts the outcomes of the subjective tests. It shows for six images (ARRI, Fly, Music, Screen, Tools, VQEG) and three codecs (JPEG XS, Tiled JPEG 2000, VC-2) the subjective quality metric given in equation (1). It reveals that both JPEG 2000 and JPEG XS clearly outperform VC-2. Moreover, except for the Fly image at 6 bpp, JPEG XS is for the tested compression ratios equal or even superior to JPEG 2000.

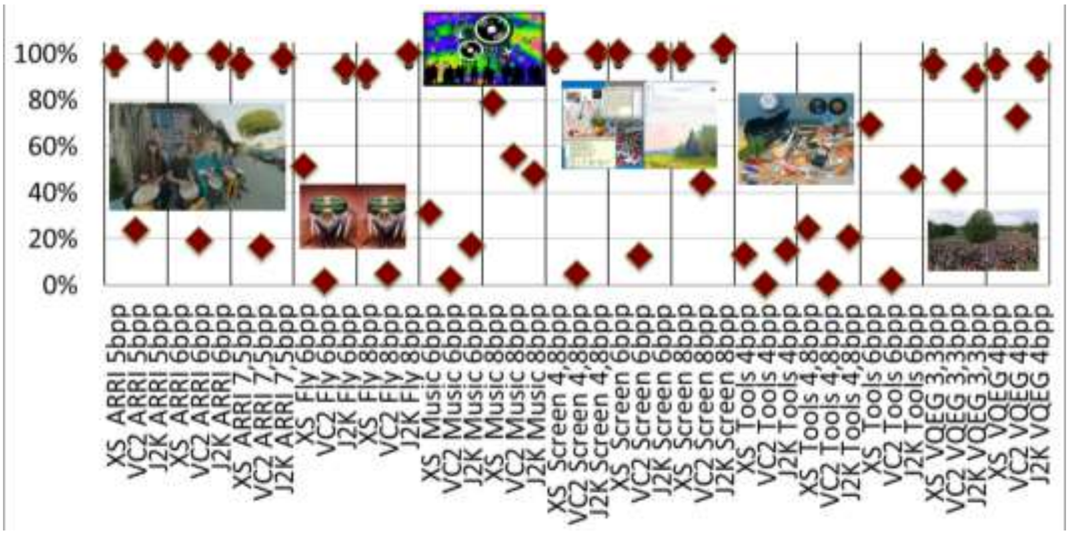


Figure 5: Results of subjective evaluation

CONCLUSION AND OUTLOOK

This paper has presented a novel low complexity video codec that aims to compensate for continuously increasing bandwidth requirements in movie and broadcast production networks. Despite the ongoing standardization process and resulting codec improvements, the quality evaluation presented in this paper already reveals a very good performance, in particular for multi-generation applications. As a consequence, the first balloted standardization document is expected for October 2017, while the final standard should be available in April 2018. In parallel, application specific profiles will be defined that restrict the permitted parameter combinations to increase interoperability. Moreover HDR support will be investigated.

REFERENCES

1. Massimo, V. and Hoffmann, H., 2008. HDTV production codec tests. EBU technical review.
2. Cronk, M. and Meyer, C., 2016. Managing Delay in Live IP Production Systems. VSF October Meeting Series. Broomfield, Colorado, USA

3. McNally, D., Bruylants, T., Willème, A., Schelkens, P., Ebrahimi, T. and Macq, B., 2017. JPEG XS call for proposals subjective evaluations. SPIE Optical Engineering + Applications. USA
4. Willème, A. and Macq, B., 2017. Overview of the JPEG XS objective evaluation procedures. SPIE Optical Engineering + Applications. USA
5. Information Technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines. ISO/IEC 10918-1 | ITU-T Recommendation T.81. 1992
6. Information technology – Lossless and near-lossless compression of continuous-tone still images – Baseline. ISO/IEC 14495-1 | ITU-T Rec. T.87. 1998
7. Information technology -- Scalable compression and coding of continuous-tone still images -- Part 1: Scalable compression and coding of continuous-tone still images. ISO/IEC 18477-1. 2015
8. Information technology -- JPEG 2000 image coding system: Core coding system. ISO/IEC 15444-1:2004 | ITU-T Rec. T.800. 2015
9. High efficiency video coding. ISO/IEC 23008-2 | ITU-T Rec. H.265. 2013
10. VC-2 Video Compression. SMPTE ST 2042-1. 2012.
11. SMPTE Engineering Project (ANSI). Apple ProRes Bitstream Syntax and Decoding Process. SMPTE registered disclosure document RDD 36. 2015.
12. Video Electronics Standards Association. VESA Display Stream Compression (DSC). VESA Standard. 2014.
13. JPEG XS Ad-Hoc Group, JPEG XS anchor configurations for evaluation process. [online] https://github.com/uclouvain/opentestbench/blob/master/doc/anchor_configurations.pdf