SMPTE Meeting Presentation

JPEG-XS - A high quality mezzanine image codec for video over IP

Thomas Richter, Joachim Keinert, Siegfried Foessel, Fraunhofer IIS Germany;

Antonin Descampe, Gaël Rouvroy, Jean-Baptiste Lorent, intoPIX Belgium

Written for presentation at the

SMPTE 2017 Annual Technical Conference & Exhibition

Abstract. *More and higher quality UHD content is arriving in the production environment, requesting additional bandwidths for data transmission and exchange. In parallel, a more flexible infrastructure based on the well-known IP protocol stack is very desirable. Adding mezzanine compression in the production workflow can reduce the necessary data transmission capacities or even enable the usage of existing infrastructure for higher resolution and higher quality content designed for previous HD production lines. A low complexity of a mezzanine codec with ultra-low latency by preserving highest quality is one of the biggest challenges for such a new codec design. Having this in mind the JPEG committee started a new work item, called JPEG-XS, addressing the need for an interoperable video-over-IP codec. This paper presents the specific requirements for such a codec, shows the results of the call for proposals, the advances during the core experiment phase and provides some insight into the selected technology.*

Keywords. Mezzanine compression, Video over IP, Low-latency

Introduction

Today, two major trends can be observed in the video production industry; One trend is to increase the resolution to 4K (UHD-1) or even 8k (UHD-2), combined with higher dynamic range (HDR). This results in a tremendous increase of image data and bandwidth requirements. The second trend is to get rid of specialized cabling and infrastructure like SDI and to use off-the-shelf infrastructure with Ethernet based information technology. The question now is: In which way can these video data streams, which are uncompressed between 3-40 GBit/s, be transmitted effectively?

A well-established method in the industry is to use compression. However, as data has to be processed within the delivery chain of a typical production facility, delays or continuous degradation due to compression-decompression cycles – so called “generation losses” – are highly undesirable for this particular application. The complexity of many existing solutions, e.g. video compression with HEVC or even JPEG 2000 compression, is also considered too high for cost-effective solutions. For this reason, the JPEG committee started in 2016 a new work on a low-complexity, low-delay mezzanine image compression codec for video-over-IP applications, and asked parties to propose relevant coding technology. The name of this new work item is “JPEG XS”, short for “extra speed” or “extra small”.

JPEG Standardization

The working group ISO/IEC JTC1 SC29 WG1, better known under the name “JPEG”, develops still image compression standards since nearly 30 years. Whereas the digital representation of compressed images was driving its initiatives in the early days, development is nowadays driven by new applications such as higher dynamic range, lower latency, and also backwards compatibility to well-deployed standards. Figure 1 shows the typical standardization process.

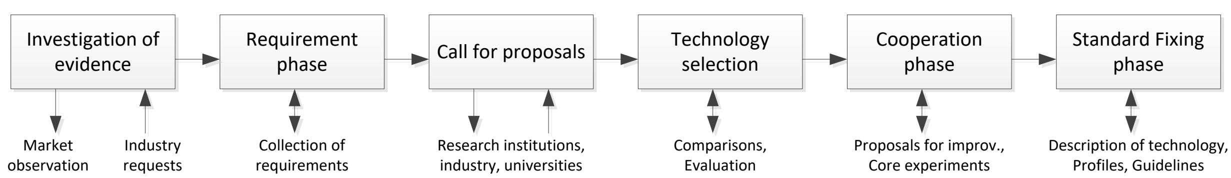


Figure 1: Simplified scheme for JPEG Standardization process

In 2016, the committee investigated potential for such new application areas and asks the industry for evidence; this resulted in the identification of the need for standardization of a mezzanine image codec. At present, multiple proprietary codecs are in use in this particular application domain, though the lack of standardization prevented the wide application of mezzanine codecs in the industry.

Requirements

In the 70th and 71st meeting, the JPEG committee identified requirements for the desired new work item. The most relevant corner points of the requirements are as follows:

- RGB/444 and YCbCr 444/422 image formats of up to 12 bits per component sample precision
- Visually lossless compression, i.e. no visible degradation
- Maximal 32 lines end-to-end (compression-decompression) latency.
- Low complexity, defined as a maximum percentage of a specific FPGA
- No external frame buffer required. In particular, individual frames shall be decoded independently
- Multi-generation robustness, i.e. minimual continuous degradation over multiple compression-decompression cycles.
- Support for multiple platforms e.g. FPGA, ASIC, GPU and CPU,
- Real-time software implementation capability for 4k/60p formats on today's standard computers

Table 2 lists some of the identified use cases, including the video format and the underyling link and its bandwidth. From this, one can estimate the required compression ratio. As seen from the table, in worst case compression ratios of up to 6:1 are necessary, which is equivalent to a bitrate of 4bpp on 24 bit RGB 444 content. For the purpose of core experiments, the JPEG committee even considers bitrates as low as 3bpp.

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

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

JPEG furthermore identified the following anchors to compare incoming proposals to the state of the art: JPEG 2000, run in a latency-constrained mode using tiles or precincts to reach the 32-line latency; the VC-2 (SMTPE 2042-1-2009/2042-2-2009) standard, based on the BBC-developed Dirac codec, addressing a very similar market. The anchors also include JPEG and HEVC, even though it does not address all requirements, especially the latency constraint for both, and the complexity constraint for the HEVC anchor.

The JPEG XS Standardization process

JPEG XS standardization started with the Request for a New Work Item as ISO/IEC 21112 at its 69th meeting, though refinement of requirements and the preparation of the Call for Proposals continued throughout the 71st meeting, following the typical layout of any ISO process as shown in figure 1. The standard was already designed to consist of multiple parts of which the first part describes the core coding system. The remainder of this paper focuses on this part, though additional parts are under preparation. They will describe profiles and buffer models in part 2, and container formats – most notably a file format and transport over SDI link – in part 3. Parts 4 and 5 traditionally define conformance testing and the reference software. Table 2 lists all currently considered parts along with the desired target dates for the publication of the DIS (draft international standard) document.

Table 2: Standardization workplan for JPEG XS (ISO/IEC 21122)

Work item	Description	Target IS date
21122-1	Part 1: Core Coding System	07/2018
21122-2	Part 2: Profiles and buffer models	07/2018
21122-3	Part 3: Transport and container formats	10/2018
21122-4	Part 4: Conformance Testing	01/2019
21122-5	Part 5: Reference Software	01/2019

Proposals

Proposal submission was divided into two phases: In the first phase, parties should indicate interest by the 72nd June meeting in Geneva, providing a broad overview on the technology they want to propose. In the second phase, ready-to-run binaries had to be submitted by September 2017 which would then enter the subjective (i.e. by human observes) and objective evaluation of the generated streams.

The following proposals were received for the Geneva meeting:

- A low-complexity wavelet based still image code for video-over-IP coding. To simplify the design, only the Haar wavelet was used here. Rate control is here run by a feedback loop that observes the output rate and steers the quantizer.

- A DCT based proposal that includes a flatness detector to avoid blocking defects in smooth image areas and a palette mode coder for screen content coding. Rate allocation operated on a line-by-line bases, with excess rate becoming available for follow-up lines.
- A modified version of VC-2 limiting the latency to the number of lines requested in the call. Similar to the first proposal and JPEG 2000, VC-2 is based on the discrete wavelet transformation.
- A codec derived from JPEG-LS which transmits a subsampled version of the image by JPEG-LS and then, based on the available rate, a residual signal to restore the full resolution.
- A modified version of JPEG 2000 replacing the high-complexity EBCOT coder by a low-complexity Mel-Coder based entropy coder that transmits multiple bitplanes at once and uses a heuristics based on frame statistics to run the rate allocation.
- A wavelet based codec that combines groups of wavelet coefficients and transmits the number of bitplanes in each group by a unary code, followed by raw transmission of the wavelet data. Unlike the first proposal, this one is based on the 5/3 filter.
- A simplified version of HEVC using only I-frame compression, and restricting the CTU-size to 8x8 blocks and the number of available prediction directions to four.

Proposals were evaluated during the 73rd meeting in Chengdu, China, including objective measurements on the basis of PSNR, and subjective evaluations following the test protocol specified in ISO/IEC 29170-2. Additionally, proponents had to provide an estimate on the complexity of their codec in terms of an approximate count in the number of lookup tables for an FPGA implementation. To test the robustness of codecs under recompression, the subjective and objective quality was also measured after 10 compression cycles. For details on the evaluation procedure, we refer to the references.

The JPEG committee considered the complexity of the JPEG 2000 based proposal and the HEVC based proposal too high to allow integration into the FPGA target architecture. Furthermore, even though the HEVC proposal showed good performance on the first compression-decompression cycle, quality degraded notably over ten generations. Despite its complexity, the JPEG 2000 based proposal performed overall quite well, though its rate-allocation is based on a frame-by-frame heuristics that causes quality drops on scene-cuts that were also considered undesirable. The VC-2 based proposal was rejected due to its bad performance compared to other submissions, even though its complexity fitted quite well to the call; the same problem could be observed on the JPEG-LS based proposal which could not reach a very high target quality over the entire test set. The first proposal, finally, was withdrawn by its proponents in favour of the second DCT-based proposal which seemed more promising.

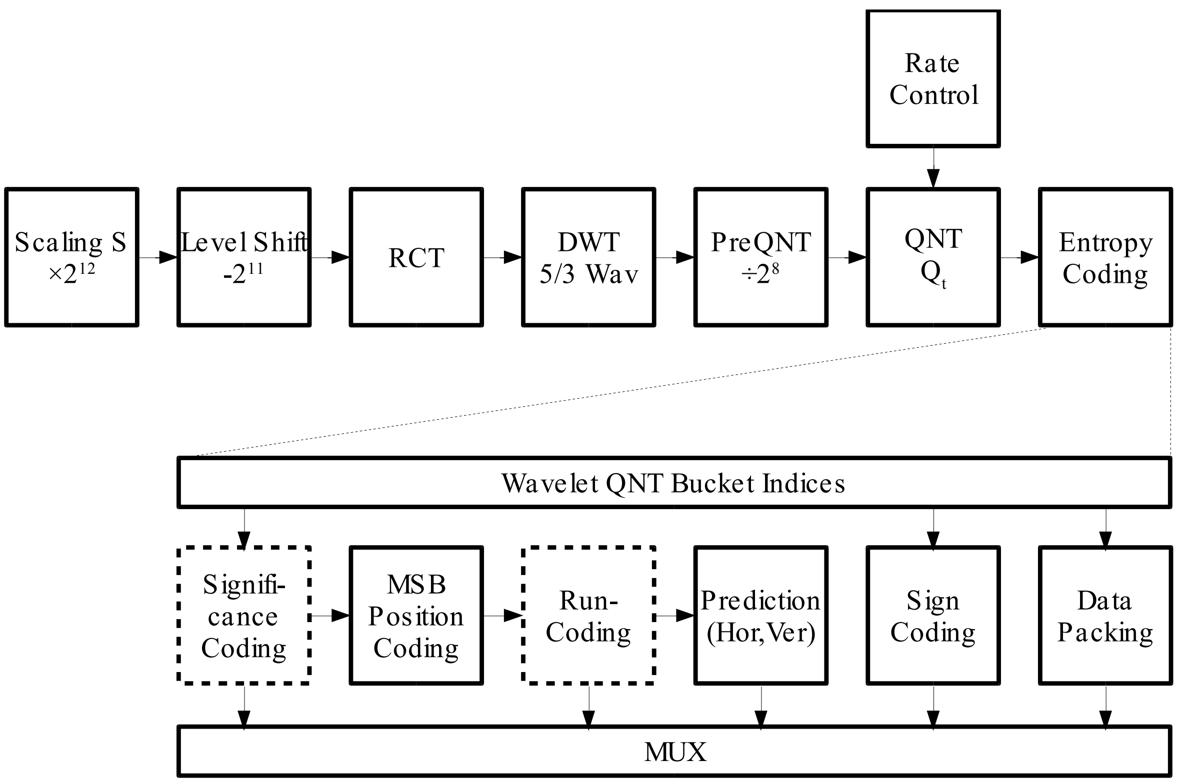
After many discussions, two proposals remained provided both consistently good performance and a complexity low enough to fit the requirements of the call, namely the DCT-based proposal – the second in the list above – and the wavelet-based technology, number 6 in the above list. It was finally decided to merge both candidate technologies into one common test model and continue standardization on the basis of this test model. A first version of the JPEG XS (“XSM”) test model became available early 2017.

Work on the combined XSM software continues throughout 2017, and multiple additional extensions and enhancements of the coding engine were proposed by the members of JPEG that have been and are under investigation by so-called “core experiments” which study the quality improvements and complexity of candidate technology. In the upcoming October meeting, the committee plans to release the first “Committee Draft” (CD) version of JPEG XS, essentially freezing the core technology and limiting ongoing development to minor improvements and corrections. The final international standard is planned to be released by the committee for publication by mid 2018.

Overview on the XSM test model encoder

Figure 2 gives a rough overview on the current JPEG XS encoder design as present in the 1.1.3 XSM test model: In the first step, input samples are upscaled, and any DC offset is removed by an appropriated shift resulting in a signal symmetric around zero. RGB components are then decorrelated by a reversible color transformation that is identical to the RCT from the JPEG 2000 standard. A 5/3 wavelet then performs energy compaction of the image signal, typically by 5 horizontal and 1 vertical transformation. This step is then followed by a pre-quantization consisting of a downshift by 8 bits, followed by regular quantization which may, depending on the rate allocation process, remove additional bitplanes. The resulting quantization bucket indices are then entropy coded.

Figure 2: Overview on the JPEG XS encoding process



Following the low-complexity design goals of JPEG XS, entropy coding is extremely simple: First, wavelet coefficients are combined into groups of 4 horizontally-adjacent coefficients each, so called "coding groups". For each coding group, the topmost populated bitplane of all four coefficients combined is computed. This bitplane count is called "MSB position" in the standard, and it is the only quantity that undergoes variable length coding:

- In the first step, significance groups of 8 coding groups, i.e. 32 coefficients, are formed, and a single bit-flag for each significance group is transmitted, indicating whether any of the coding groups in the group carries data at all. If the significance flag is set, the entire significance group is skipped and all 32 coefficients within the group are assumed to be zero by the decoder.
- In the second step, the MSB positions of all remaining significant coding groups are transmitted. JPEG XS provides multiple coding tools for MSB position coding: Either, MSB positions are transmitted in raw, 4 bits per coding group. Or MSB positions are predicted from their left or top neighbor, and the prediction residual is transmitted in a unary code. Optionally, a run-length code may skip over runs of empty coding groups.
- Third, for all significant coding groups, the absolute value of the quantized wavelet coefficient is transmitted. This includes all bits from the MSB position down to the bitplane the quantizer selected.
- Fourth, for all non-zero coefficients, the sign bits are transmitted directly.

The following sections shed additional light on the energy compaction and rate allocation, and how JPEG XS requirements influenced the selection of the transformation and the rate allocation strategy.

Design constraints on energy compaction and rate allocation

The requirement of low-latency has several implications on the design of JPEG XS, most notably on the choice of the decorrelation transformation and on rate allocation. The purpose of the former is energy compaction, i.e. the image data should be represented by only a few coefficients; rate allocation then has to adjust the information loss – typically by a quantizer – to generate a codestream that fits to the bandwidth requirements of the output channel. Redundance reduction, i.e. entropy coding, plays only a minor role in the JPEG XS standard as the desired overall compression ratios are relatively moderate. The design of JPEG XS is hence mostly a rate-allocation problem, together with the choice of a suitable energy compaction.

The maximum end-to-end latency of 32 lines leaves at best 16 lines latency for the energy compaction at the encoder, i.e. decorrelation transformation. Under realistic conditions, additional lines are required for rate allocation, buffering and codestream build-up.

Now, in all proposals received, the following strategies could be observed for energy compaction:

- Spatial-domain strategies based on prediction schemes. The JPEG-LS variant can be classified as such an approach, though subsampling and transmission of a residual can be considered as an application of a Haar wavelet or a 2x2 DCT. This limits latency due to energy compaction to two lines, though as experiments showed, the overall performance of such methods does not seem to be sufficient.

- Strategies based on the DCT, as used by two proposals: The second proposal which combines it with a palette mode and limits the DCT size to 4x4 blocks to keep the complexity low, and the derived/restricted HEVC variant which deploys an 8x8 DCT. While the DCT is also necessarily more complex, most reservations were due to the complexity of the HEVC backend. As demonstrated by the second proposal, the DCT can be made to meet the target requirements, though additional precautions are necessary to avoid blocking defects in low-rate situations. More on this can be found in the references.
- Strategies based on the DWT (discrete wavelet proposals), deployed by all other proposals, including variants of JPEG 2000, VC-2. What is common to all submissions is that the vertical wavelet filter has to be relatively short, and only a very limited number of vertical wavelet decompositions is possible, see Table 3. As listed there, a 5/3 wavelet can be run in one or two decomposition levels, though the longer 13/7 wavelet can only operate in a single vertical decomposition level. There is no restriction on the number of horizontal wavelet decompositions; received proposals run typically five horizontal decompositions. Wavelet-based approaches have demonstrated that they are able tor meet the quality and complexity requirements of JPEG XS. The impact of the wavelet filter on the coding efficiency will be studied in this paper as well, and we will present results below.

Table 3: Latency of various wavelet filters under consideration

	5/3 wavelet, 1 level	**5/3 wavelet, 2 levels**	**13/7 wavelet, 1 level**
Lo-pass, encoder	2	6	6
Hi-Pass, encoder	1	4	3
Lo-pass, encoder	1	3	3
Hi-pass, decoder	2	5	6
End-to-end delay	3	9	9

Rate control is the second critical component of JPEG XS. Naturally, the more image data is available for rate-allocation purposes, the more precise it can operate, and the better the image quality may become.

- Some strategies proposed were based on heuristics, using the statistics of the previous frame to allocate the rate for the current frame. For example, the JPEG 2000 based proposal could not use a full EBCOT tier 2 rate allocation due to the latency constraint imposed by the JPEG XS requirements. Any heuistic, however, has the drawback that quality degradations may appear on scene cuts, i.e. under dramatic changes of the coefficient statistics. While such degradations are masked away by the scene cut and are thus typically not visible for human observers, they may be still undesirable in applications where material is post-processed after transmission. They would introduce an additional error source that should be avoided.
- Other proposals used a line-by-line rate-allocation and make excess rate not spend in the current line available to the next if bandwidth and latency permits. For DCT based

proposals, rate allocation included in one proposal a detection of blocking defects and assigns more rate to blocks that would be succeptible to such defects.

- Another approach is to run rate-allocation on a somewhat larger lookahead window, thus having the advantage of already allocating rate on the current line for critical content in the future of the encoder. Clearly, the size of the window cannot exceed the maximum tolerable latency as the encoder has to delay its coding process until the full lookahead window becomes available.
- Finally, some proposals used a rate-control-loop approach that observes rate spend so far, adapting quantization as necessary. The control-loop approach was refined in one proposal by a model of human vision such that quantization defects remain below the just-noticable difference.

Overall, experiments have shown that the approach based on a lookahead window works best, though experiments on alternative wavelet filters such as the 13/7 filter are still ongoing. Preliminary results will be presented below. While longer filters provide better quality, they also reduce the amount of lookahead available for the rate-allocator. Finding a good trade-off between the latency taken by rate-allocation and the latency introduced by energy-compaction is the purpose of the ongoing work of the committee.

Parallel processing

While low-latency is one aspect of JPEG XS, parallel processing is another: GPU and vectorized CPU implementations are one important application domain for the envisioned standard. While the wavelet transformation and the color decorrelation transformation is trivial to parallelize, entropy coding requires additional care. Multiple design choices allow parallel decoding:

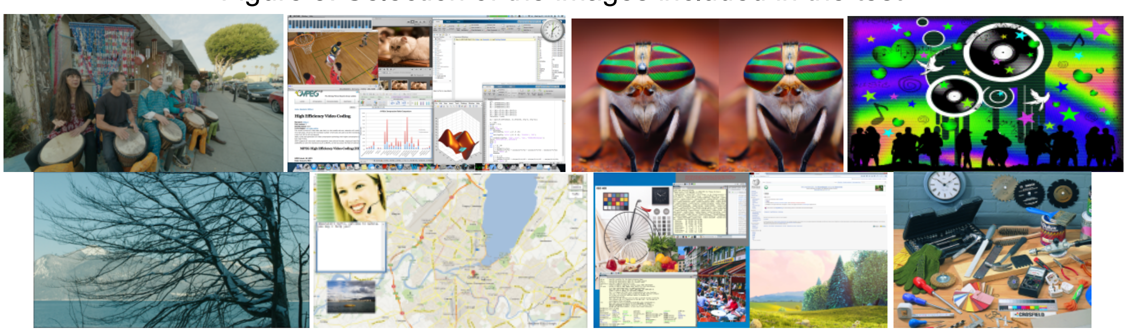
- Combining coefficients into coding groups helps vectorizing the decoding on CPUs equipped with modern vector instructions.
- Of the four entropy coding phases discussed in the section above, only a single data item is encoded at all, namely the MSB positions of the coding groups. All other elements are directly copied into the bitstream. MSB position coding uses an extremely simple alphabet, namely unary coding. This code transmits a positive value N by N 1-bits followed by a 0 comma bit. A unary code can be decoded in parallel as it is self-synchronizing: By starting at an arbitrary position within a unary-encoded bitstream, at most the initial symbol may be decoded incorrectly, but all follow-up symbols will be decoded correctly. By running multiple decoders on slightly overlapping blocks, this observation can be used to design a fully-parallel unary-alphabet entropy decoder. Finally, horizontal and vertical prediction are again relatively easy to parallelize.

Experiments and results

To evaluate and demonstrate the performance of the JPEG XS coding system and provide a more complete picture on the technology available on the market, we measured the rate-distortion curves of multiple image still image codecs that address related use cases. In particular, we measured the following codecs and configurations:

- JPEG (ISO/IEC 10918-1), represented by the University Stuttgart JPEG XT Reference software. The codec is configured for 444 input, optimized Huffman coding and Trellis quantizer optimizations. Note that the tested configuration is tuned towards optimal visual quality, not optimal PSNR. Furthermore, JPEG does not guarantee a latency, nor a budget constraint. Its main purpose in this test, as in the core experiments, is to provide an anchor.
- JPEG 2000 (ISO/IEC 15444-1) in a low-latency configuration. In this particular mode, each frame is segmented in stripes of height pixels height, where each stripe is coded independently. This configuration is a typical broadcast application, limits the latency and the bandwidth, but does not provide ideal rate/distortion performance. In particular, the implementation under test is here the Accusoft software as it implements the required mode.
- VC-2 (SMPTE 2042-1-2009/2042-2-2009) as a wavelet-based still-image codec that was especially developed for broadcast applications. A variant of VC-2 also participated in the JPEG internal core experiments. It is here configured to use the 5/3 wavelet (same as JPEG XS) with three wavelet decompositions and 1x2 slices, in the low-delay configuration. For testing, the VC-2 reference implementation has been deployed. Images are first upscaled to 12 bit, then converted to YCbCr and finally compressed with VC2.
- Apple ProRes, a DCT-based mezzanine codec also targeting the broadcast market. Similar to JPEG, it does not offer latency- or bandwidth constrained coding, and offers only five separate configurations of (only approximate) target bitrates. In this work, we run ProRes in all available configurations and measure the rate and distortion in each possible combination. Each configuration enters as one points in the plot, giving it a somewhat less regular shape. Input and output is converted to and from 16 bit 422 YUV if necessary, PSNR is always measured after converting the reconstructed image in 444 RGB.

Figure 3: Selection of the images included in the test



From left to right, top to bottom: AlexaDrums, a natural scene. AppleBaseball: Mixed natural and screen content, FemaleHorseFly: Smooth gradients with highly textured image regions, Hintergrundmusik: Highly textured artificial content, Lake: A natural image with flat regions and highly structured areas, HuaweiMap: A screen capture containing a rendered street map, RichterScreenContent: Mixed raster content combining text and natural images, Tools: An oversharped image from the ITU test set.

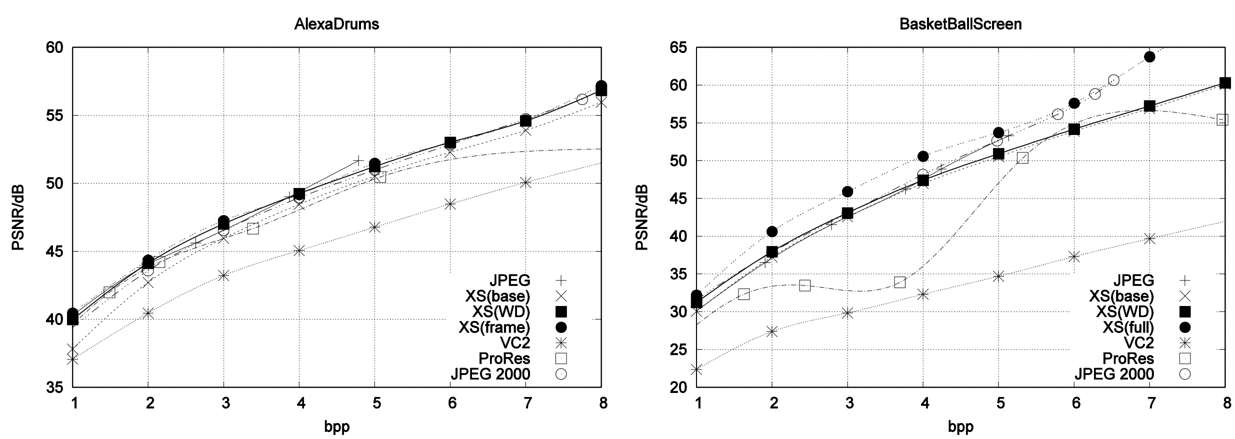
- JPEG XS, in three configurations: The initial configuration is identical to what the initial 1.0.0 XSM software offered, namely the 5/3 wavelet with a single vertical and five horizontal decomposition levels, with horizontal or vertical prediction, sign and significance coding. This corresponds to the features that became available with the first (1.0) XSM software.

- A second, more advanced configuration of JPEG XS which adds features such as a second vertical decomposition level and additional advanced entropy coding modes of the MSB positions such as tag-tree based modes. This corresponds to the XSM version 1.1.3 and coding tools within may become available at the CD stange of JPEG XS at the 77th meeting.
- A third version of JPEG XS that does not include any latency constraint and a full-frame rate-allocation. Otherwise, this third configuration is identical to the second.

Source material was taken from the JPEG XS core experiments test set, though data was converted and downscaled to 8bpp RGB 444 for all codecs, in particular to allow compression with JPEG. For Apple ProRes, a configuration-dependent conversion from 444 RGB to 422 YUV was included in the experiment, and each data point in the plot corresponds to one of the overall five possible configurations of the codec. Unfortunately, ProRes does not allow any form of rate control beyond this very coarse adjustment. PSNR was measured, however, consistently in 444 RGB space. Figure 3 shows the images that were included in this test.

Figure 4 shows the rate distortion performance of all selected codecs on the source material listed in Figure 3. PSNR is here plotted over bitrate. Note that the scale varies from image to image as performance depends on the complexity of the source.

The rate-distortion performance of the multiple wavelet filters available is plotted in figure 5. The entropy coding tools were here limited to the set available in WD1.0, i.e. no advanced coding techniques were deployed. However, the wavelet filters were varied between the 5/3 filter, the 5/3 filter with two instead of one vertical decomposition level and the 13/7 wavelet with a single vertical decomposition level. The rate allocation window and the rate allocator lookahead were configured to be as large as possible while still ensuring a latency not larger than 32 lines. In particular, this allows a larger lookahead for the simple single-level 5/3 filter, followed by the 13/7 filter and the 2-level 5/3 filter which allows the smallest lookahead of all. Even though the latency of the 13/7 and the 2-level 5/3 filter are exactly identical, the latter allows only rate allocation in groups of four lines, unlike the single level 13/7 filter, resulting in an additional restriction on the lookahead window size for the two-level filter.



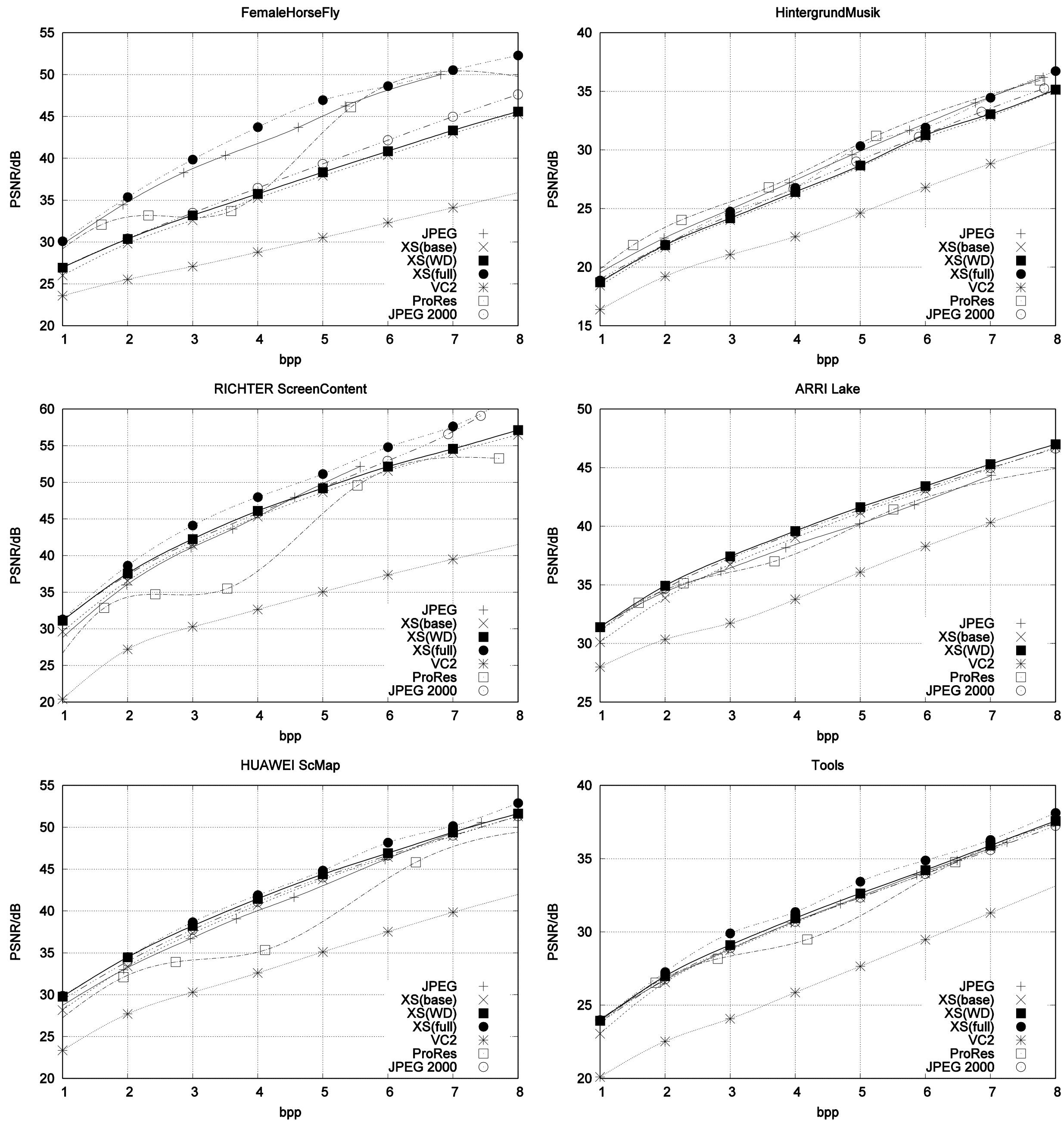
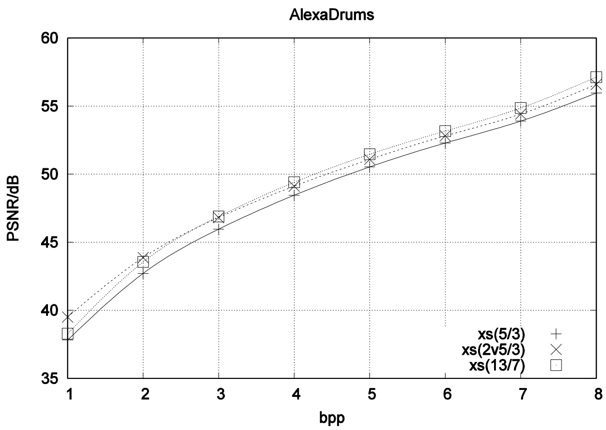
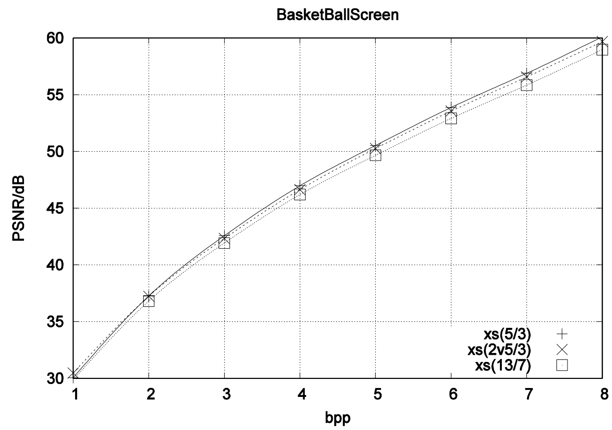
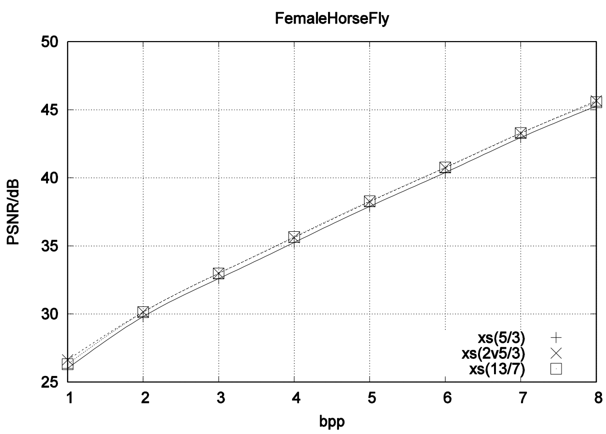
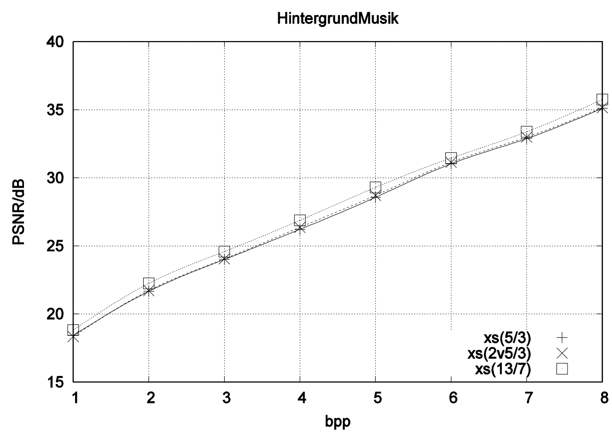


Figure 4: Rate/Distortion performance of all tested codecs on the images in Figure 3.

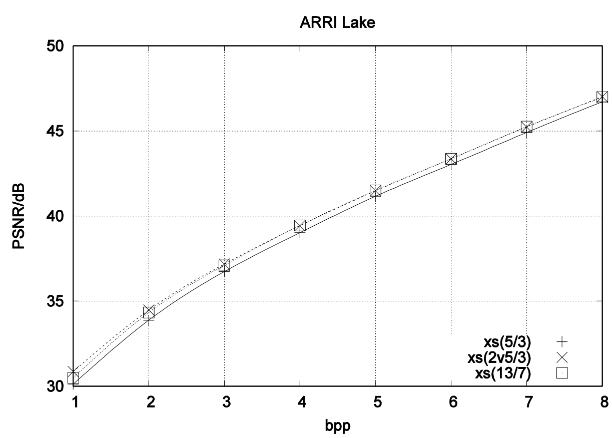

AlexaDrums
60
55
50
45
40
35
PSNR/dB
1 2 3 4 5 6 7 8
bpp
xs(5/3)
xs(2v5/3)
xs(13/7)


BasketBallScreen
60
55
50
45
40
35
30
PSNR/dB
1 2 3 4 5 6 7 8
bpp
xs(5/3)
xs(2v5/3)
xs(13/7)


FemaleHorseFly
50
45
40
35
30
25
PSNR/dB
1 2 3 4 5 6 7 8
bpp
xs(5/3)
xs(2v5/3)
xs(13/7)


HintergrundMusik
40
35
30
25
20
15
PSNR/dB
1 2 3 4 5 6 7 8
bpp
xs(5/3)
xs(2v5/3)
xs(13/7)


RICHTER ScreenContent
60
55
50
45
40
35
30
PSNR/dB
1 2 3 4 5 6 7 8
bpp
xs(5/3)
xs(2v5/3)
xs(13/7)


ARRI Lake
50
45
40
35
30
PSNR/dB
1 2 3 4 5 6 7 8
bpp
xs(5/3)
xs(2v5/3)
xs(13/7)

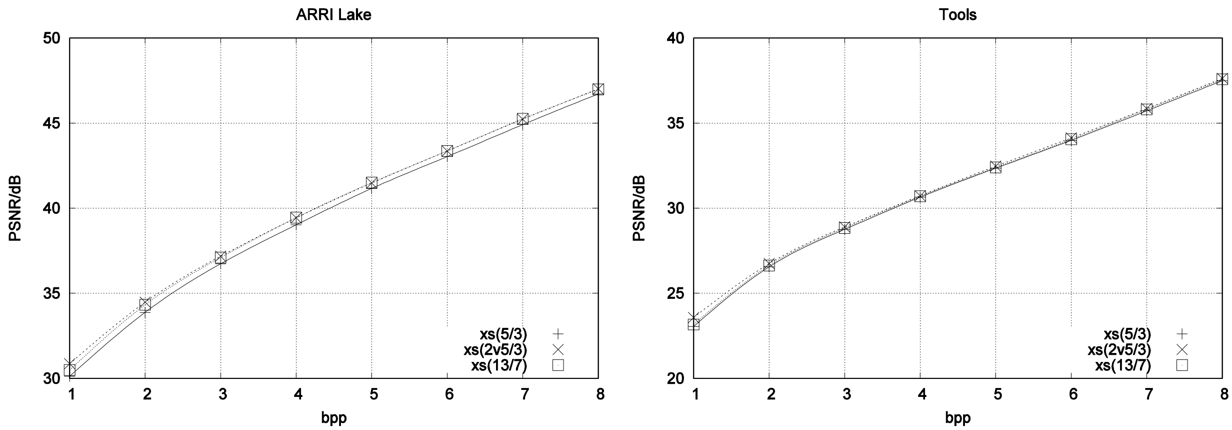


Figure 5: Rate/Distortion performance for various choices of the wavelet.

Discussion

Depending on the content, JPEG XS is on par or even outperforms the low-latency variant of JPEG 2000. It works particularly well on natural and smooth content, though DCT based methods like JPEG and Apple ProRes show an advantage on complex content. Note, however, that both codecs do not include a rate control and hence do not address the same set of requirements as JPEG XS. Given the simplicity of the codec, results are very promising and show competitive coding performance at bitrates relevant for the use cases listed in Table 1.

Comparing the latest version of JPEG XS as found in the August 2017 working draft 4.1 / XSM version 1.1.2 with the initial version 1.0.0 from late 2016 shows a significant improvement of coding performance of around 2dB on average. This is mostly due to the 2 vertical level wavelet decomposition, but additional coding tools such as runlength coding also help to improve the coding gain.

Complex images such as the “FemaleHorseFly” image show the limitations of latency-constrained coding: Codecs such as JPEG or ProRes perform considerably better here than latency-limited codecs such as JPEG XS or the constrained JPEG 2000 variant used for testing. Releasing this constraint improves coding performance considerably as direct comparison of various JPEG XS configuration demonstrates. The tested “full frame” variant tested here is, however, beyond the requirements of the JPEG XS standard.

Figure 5 shows that the choice of a two-level wavelet decomposition improves rate/distortion performance by up to 3dB for very low bitrates, which are again outside of the JPEG XS requirements. The performance of the 13/7 filter is somewhere between the 1 and 2 level 5/3 filter. For bitrates within the requirements of the standard, the 13/7 filter reaches approximately the performance of the 2-level filter, and in one case – namely the very complex “Hintergrundmusik” image – outperforms all other filters. However, it is at this point still unclear whether this filter will be included in the standard or not.

The JPEG XS standard targets even more moderate compression factors up to 2:1 which are not shown in the graphs generated for this work. We did not observe any particular change in the behavour beyond 5bpp, and the graphs can be extrapolated beyond 8bpp without changing the relative performance of the codecs to each other.

Conclusion

The new JPEG XS standard offers an optimized solution for a mezzanine compression codec, where the design space was explored to allow a very high compression quality in applications targeting a compression factor from 2:1 to 6:1 with minimal resource consumption and low-latency. According to the time plan, the JPEG standard will be finalized in 2018, and first implementations will be expected in 2018.

Acknowledgements

The authors want to thank the JPEG Committee for providing information on the requirements and evaluation results.

References

ISO/IEC JTC1 SC29 WG1, "JPEG XS - Call for Proposals for a low-latency lightweight image coding system", https://jpeg.org/items/20160311_cfp_xs.html, March 11, 2016.

ISO/IEC JTC1 SC29 WG1, "Information technology — Advanced image coding and evaluation — Part 2: Evaluation procedure for nearly lossless coding", https://www.iso.org/standard/66094.html , August 2015.

S. Daly, "The visible difference predictor: an algorithm for the assessment of image fidality", in Digital images and human vision, pp. 179-206, MIT Press Cambridge, MA, USA, 1993.

R. Mantiuk, K-J Kim, A.G. Rempel, W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions", ACM Transactions on Graphics (proc of SIGGRAPH'11), 30(4), 2011

Th. Richter, S. Fößel, J. Keinert, A. Descampe, "High-Speed, Low-Complexity Image Coding for IP Transport with JPEG XS", Proc. of Appl. Of Digital Image Proc. XXXIX (SPIE 2016), 2016

A. Descampe, J. Keinert, Th. Richter, S. Fößel, G. Rouvroy, "JPEG XS: a new standard for visually lossless low-latency lightweight image compression", Proc. of Appl. Of Digital Image Proc. XL (SPIE 2017), 2017

A. Willème, Th. Richter, C. Rosewarne, B.M. Macq, "Overview on the JPEG XS objective evaluation procedures", Proc. of Appl. Of Digital Image Proc. XL (SPIE 2017), 2017

D. McNally, T. Bruylants, A. Willème, P. Schelkens, T. Ebrahimi, B.M. Macq, "JPEG XS call for proposals subjective evaluations", Proc. of Appl. Of Digital Image Proc. (SPIE 2017), 2017

Th. Richter, J. Keinert, S. Fößel, "High-speed low-complexity video coding with EDiCTius: a DCT coding proposal for JPEG XS", Proc. of Appl. Of Digital Image Proc. XL (SPIE 2017), 2017

