ISO/IEC JTC 1/SC 29/WG 1
(& ITU-T SG16)

Coding of Still Pictures

JBIG	**JPEG**
Joint Bi-level Image Experts Group	Joint Photographic Experts Group

TITLE:	Use Cases and Requirements for JPEG XS v3.2
SOURCE:	WG1 – Requirements subgroup
EDITORS:	Tim Bruylants (t.bruylants@intopix.com) Thomas Richter (thomas.richter@iis.fraunhofer.de)
PROJECT:	ISO/IEC 21122 (JPEG XS)
STATUS:	**Final**
REQUESTED ACTION:	Publish
DISTRIBUTION:	Public
NOTE:	

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

Contents

Abstract

Image sequences have been transmitted and stored in uncompressed form in many cases, such as in professional video links (3G/6G/12G-SDI), IP transport (SMPTE ST 2022-5/6, SMPTE ST 2110-20, and proprietary uncompressed RTPs), Ethernet transport (IEEE/AVB), and memory buffers. A low-latency lightweight image coding system allows for increased resolutions and frame rates, while offering visually lossless quality at minimum, under reduced resource constraints such as power and bandwidth.

This document provides the most current use cases and requirements for such a low-latency lightweight image coding system.

Introduction

Infrastructure and system capabilities increase at a slower pace than content resolution and are therefore steadily becoming the bottleneck in many applications.

As an example, the Olympic Broadcasting Service (OBS) started initial production of 8K content during the 2020 Tokyo Olympics and will upscale this for the upcoming 2024 Paris Olympics. Transmitting uncompressed 8K content over existing links or over soon-to-be-available alternatives channels is not feasible, especially at high frame rates such as 60 or even 120 fps.

In addition, other technologies such as 4K/8K (UHD) HDR, live immersive 180° stereoscopic and 360° panoramic coverage, are also on the rise, all suffering from similar bandwidth limitations that need to be addressed.

In combination with the ongoing transition to cloud-based production, it becomes evident that a lightweight low-latency coding system is needed to offer a smart and affordable solution to meet market demands.

Even for transmitting content that fits currently available systems, the use of a lightweight and low-latency coding system can still be beneficial as it allows for reducing bandwidth, and consequently results in lowering corresponding cost or enable longer cable runs (for example, usual 3G-SDI cable run is 14m while it reaches 21m for HD-SDI, and 50m for SD-SDI), and also enabling transmission over 5G networks.

In a nutshell, the key advantage of a lightweight and low-latency image coding system is to allow increasing resolution, component depth and frame rate in a cost-effective manner, i.e.:

- **Safeguarding all advantages of an uncompressed stream**
  - low power consumption (through lightweight image processing).
  - low latency in coding and decoding.
  - easy to implement (through low complexity algorithm).
  - small size on chip and fast software running on general purpose CPU with the use of SIMD and GPU.

- **Reducing the required bandwidth**
  - low power consumption (through reasonable bandwidth interfaces).
  - longer cable runs.
  - SRAM size & frequency reduction with a frame buffer compression.
  - more adequate for current infrastructures.

JPEG XS Use Cases and Requirements v3.2

1 Scope

This document presents the JPEG XS use cases and requirements. This version specifically adds 5G related use cases and associated requirements.

The primary goal of JPEG XS is to provide a low-latency lightweight image coding system that can be used in the following markets:

- Broadcast applications and live production,
- Live production,
- Digital Cinema applications,
- Industrial vision,
- Professional audio-visual systems,
- Consumer TV,
- Mobile video applications,
- Camera array-based recordings,
- Ultra-high frame rate cameras,
- Medical Imaging,
- Video Surveillance and security,
- Automotive Infotainment,
- Camera manufacturers,
- Set-top boxes,
- Low-cost visual sensors in Internet of Things (IoT),
- HMD displays,
- Video streaming over 5G networks.

2 Key features

This section provides an overview of the key features that were identified to be important for JPEG XS, explaining the general idea of how JPEG XS differentiates itself from other existing image coding systems.

2.1 Implementation architecture

It is important for JPEG XS to allow efficient implementation on various platform architectures – classified here into CPU, GPU, FPGA and ASIC – without impairing the compression performance and without imposing any platform dependent algorithmic choices. The platform architecture that generated a codestream should not have any influence on the outcome of the encoding process, and vice versa. As such, encoding and decoding can happen on any architecture independently, making JPEG XS a cross-platform technology. Furthermore, each platform architecture has its own benefits and drawbacks. In this respect, the algorithm of JPEG XS must support maximal utilization of each platform architecture's resources.

For **CPU** implementations, this means:

- Allow maximal usage of optimized/vector/specialization instructions on common architectures (e.g. SIMD, AVX, etc.).
- Natively facilitate parallelization at both the instruction level and the threading level (i.e. small and large parallelization granularity).
- Prevent branching as much as possible (improving for pipelined instructions, prefetching, execution prediction, etc.).
- Minimize external memory requirements. On typical CPU platforms, relying on external memory is not a huge issue, but less is better.

- Monitor cache memory constraints and capabilities.

For **FPGA** implementations, this means:

- Limit reliance on internal block memory, which is a costly and constraint resource on many FPGAs.
- The processing pipeline should be irrespective of the pixel-per-clock-cycle, by preventing data dependencies, allowing multiple clock domains (fast-clock vs slower-clock) and enabling computational parallelization.
- Full utilization of the extra computational resources on an FPGA allowing for new and better coding tools.
- External memory may be allowed, provided that memory bandwidth is under control (e.g. using a compressed frame buffer).

For **ASIC** implementations, this means:

- Support exploitation of the faster clocks (compared to FPGA) which means less pixels per clock and allows for reduction of the footprint.
- Limit reliance on internal memory.
- External memory may be allowed, similarly to the FPGA case.

And, finally, for **GPU** implementations, this means:

- Make optimal use of large amounts of external memory, typically available on GPU systems. Yet, also account for the limited memory bandwidth by keeping it under control.
- Exploit the high degree of parallelization available on a GPU. The typical computation capability is between CPU and FPGA/ASIC.

2.2 Low latency

In principle, JPEG XS strives to keep the incurred encoding-decoding latency to an absolute minimum, in particular below a fraction of a video frame. However, in some applications having extreme low latency is a less hard requirement, thus relaxing this feature for certain use cases should be possible. Allowing a slightly larger latency provides more opportunity to improve visual quality or lower the computational-vs-memory requirements. The JPEG XS standard should ideally list use cases and their respective allowed maximum latency and make a classification into latency classes/profiles.

2.3 Bit rate allocation

Most use cases, especially when low latency is required, will rely on constant bit rate (CBR) allocation, to ensure that data can be transferred over limited bandwidth channels. Yet, in some cases it can be beneficial to also allow a form of variable bit rate (VBR) allocation. Examples are mathematically lossless (MLS) coding or scenarios where latency can be relaxed a bit, allowing more time to perform rate-distortion optimization to improve the visual quality.

2.4 Visual quality

The main objective of JPEG XS is to achieve visual transparent quality with compression bit rates as low as 3 bits per pixel (bpp). This means that compressed content is indistinguishable from the original content (visually lossless quality) or that it does not impair machine vision processing (automatic-analysis-resilient quality). Moreover, in a compressed video sequence, any compressed image frame from the sequence shall individually achieve such quality.

However, the absence of a reference at the receiver side may allow higher compression ratios in the high to near-visually lossless quality range as defined in the JPEG AIC-3 project [9]. Such minimal visual degradation of the image quality will allow for significant savings in the required bandwidth and can reduce cost or resources.

On the other hand, certain use cases require higher quality which shall also be achievable by JPEG XS:

- Mathematically lossless quality.

- Flickering-resilient visually lossless picture quality. In this case, the picture quality achieved after compression and decompression shall successfully pass the flickering test, as described in Annex B of ISO/IEC 29170-2:2015 [4].

In addition, JPEG XS is required to minimize the generational loss when recompressing content multiple times over.

2.5 High dynamic range (HDR) content

Compression of high dynamic range (HDR) content requires not only the support for more than 10 bits per channel, but also requires proper support for signaling properties like the color primaries, the applied transfer function, potentially floating-point sample values, etc. In addition, efficient compression of HDR content requires some specific adjustments to the compression scheme when compared to SDR content. Given the overall adoption of HDR in all relevant markets (e.g. OpenEXR [8]), JPEG XS must be able to efficiently transport HDR content.

3 Use cases

3.1 Transport over video links and IP networks

In these use cases, a video link and transport protocol are employed to transport video streams at a higher throughput than its physical throughput, thanks to a lightweight compression with a compression ratio ranging from 2:1 to 15:1. Several examples are given in the table below.

Table 1 — Example of bandwidth requirements and compression ratios for various commonly used video formats

Video stream	Video throughput [1]	Physical link	Available throughput	Comp. ratio
2K@60p 422 10-bit	2.7 Gbps	HD-SDI	1.33 Gbps	~2
2K@120p 422 10-bit	5.4 Gbps	HD-SDI	1.33 Gbps	~4
4K@60p 422 10-bit	10.8 Gbps	3G-SDI	2.65 Gbps	~4
2K@60p 422 10-bit	2.7 Gbps	1G Ethernet (SMPTE 2022 1/2)	0.85 Gbps	~3
2K@60p 444 12-bit	4.8 Gbps	1G Ethernet (SMPTE 2022 6)	0.85 Gbps	~6
4K@60p 422 10-bit	10.8 Gbps	1G Ethernet	0.75 Gbps	~15
8K@60p 420 10-bit	29.9 Gbps	2.5G Ethernet	2 Gbps	~15
4K@60p 422 10-bit	10.8 Gbps	10G Ethernet (SMPTE 2022 6)	8.5 Gbps	~1.3
Triple 4K@60p 422 10-bit	32.4 Gbps	10G Ethernet (SMPTE 2022 1/2)	7.96 Gbps	~4
4K@60p 444 12-bit	19 Gbps	10G Ethernet (SMPTE 2022 1/2)	8.5 Gbps	~2.2
Dual 4K@60p 444 12-bit	37.9 Gbps	10G Ethernet (SMPTE 2022 6)	7.96 Gbps	~5
8K@120p 422 10-bit	85 Gbps	25G Ethernet	21.25 Gbps	~4

[1] Gbps = Gigabit per second

Full support for HDR content is also becoming increasingly important, hence proper signaling support is needed (e.g. image parameter values as given in ITU-R Rec. BT.2020, ITU-R Rec. BT.2100, etc.).
This use case can be further divided into more specialized, but related, use cases.

3.1.1 ProAV video over IP

This use case targets the professional audio visual systems market and therefore puts slightly more emphasis on the overall compression performance (or quality) and less on the extreme low latencies or low complexity. Given its professional context, this use case also assumes a reliable and well-designed network (professional LAN and professional WAN) for transport (like 1 Gbps, 2.5 Gbps and 10 Gbps using ST 2110 or MPEG2-TS). Having access to slightly increased computation/memory resources and relaxing the latency allows improving the overall compression performance. Strict CBR allocation is also less important, so this use case would benefit from allowing a limited or capped form of VBR allocation to reduce bandwidth waste (e.g. due to padding).
This use case aims to improve visual quality compared to JPEG XS with the high profiles.

- Further reduce the bit rate while moderately increasing the allowed complexity, while keeping at least the same quality as with XS high profiles.
- Allow frame quality degradation on scene cuts, with the lowest frame quality bound to the quality achievable with XS high profiles at the same bit rate.
- Relax constraints regarding the flickering test of AIC-2 (ISO/IEC 29170-2:2015) [4], using a to-be-designed no-reference visual acceptance test.
- Allow spatial region decoding from a single codestream source (e.g. video wall where one encoder drives multiple layouts); Examples of targeted compression ratios in the 1 to 1.5 bpp range:
  - 4K@60p-422: 500-750 Mbps,
  - 8K@60p-422: 2 Gbps.

3.1.2 Broadcast and remote production

The broadcast and remote production use case is similar to the ProAV use case, but here JPEG XS is typically used as an intermediate transport compression format. Often the video content originates from or is destined to become encoded in an alternative compression technology. This means that efficient transcoding to and from JPEG XS while preventing quality loss becomes key.

3.1.3 Movie production

The movie production use case adds the fact that in many situations video content is stored as a master format on one device and transferred to another. This requires JPEG XS, here employed as a mezzanine codec, to be able to efficiently represent whatever video content is processed. This means also that transcoding with minimal loss of quality to and from JPEG XS is important (like JPEG 2000, ProRes etc). Moreover, in movie production samples can be represented as floating-point data to represent HDR content, which requires special care in JPEG XS to properly compress.

3.1.4 Display over IP

This use case targets applications like KVM switches (keyboard-video-mouse), remote desktop, computer display extension and similar. The extra difficulty is that the content is often artificially generated, which has significantly different properties compared to typical natural content (think CGI, screen content and GUI elements). This use case is also significantly different from the ProAV use case, because of the following properties:

- mainly using 4:4:4 content,
- high/variable frame rate,
- low latency,
- efficiently handle mixed content (GUI, CGI, text, natural content, etc),

- uncontrolled network (less reliable, more jitter and loss) such as consumer LAN, WIFI and Internet,
- smaller available bandwidth (way below 1 Gbps, and often even below 100 Mbps).

Given these properties and to be usable, the overall bit rate should be significantly reduced over what JPEG XS 2nd edition can deliver.

3.2 Real-time video storage and playout

Embedded devices such as cameras use internal storage to store large streams of images. These devices typically offer limited sustainable writing speeds (e.g. approximately 4 Gbps for SATA SSD drives or 400-720 Mbps for typical SD cards), even when adopting faster link technologies such as SD 8.0, CFExpress and NVMe. Lightweight compression would allow real-time storage of video streams with throughputs higher than these sustained writing speeds. Moreover, compression significantly increases the usable recording time for the same amount of storage capacity. In addition, efficient frame-based random access can be beneficial for playout of large streams stored on devices with limited access rates.

3.3 Video memory buffer

Buffer compression reduces the system form factor's weight, decreases the number of interconnect wires and extends the battery life for battery powered systems:

- upscaler/downscaler,
- buffer for high refresh rate displays (120~600 Hz, Triple Flash),
- storage and replay buffer for high-speed camera,
- key frame buffer for AVC/HEVC 4K decoder,
- efficient spatial random access.

3.4 Omnidirectional video capture system

Omnidirectional video capture systems are assembled from a multitude of cameras mounted on a platform. Each camera covers a certain field of view which tends to overlap with that of its adjacent cameras to facilitate image stitching.

The proposed use case addresses the challenge of concurrently transferring and storing the image streams from each camera to a front-end processing system. In order to reduce the required bandwidth and therefore allow multiple cameras to send their data over a shared physical link, a lightweight, real-time compression of the image data at the camera is desired. Furthermore, this compression should be configurable in a transparent manner. Applying such compression will furthermore reduce both the required storage size and throughput demands of the storage subsystem on the front-end processing system.

3.5 Head-mounted display for Virtual or Augmented Reality (VR/AR)

Omnidirectional VR and AR content is highly suitable for viewing through Head-mounted Displays (HMD). HMDs are either tethered (i.e. connected through a cable) or wireless in which case the display is battery powered. Furthermore, with omnidirectional content, the HMD will only show that portion of the media stream which is within the viewer's field of vision. Given the computational (and power) constraints of such a display, it cannot be expected to receive the full image stream and then locally perform the required filtering onto the viewer's field of vision – this needs to be done upstream and based on spatial data received from the HMD.

From the viewer's perspective, the quality of experience is crucially tied to the latency with which the system reacts to changes in his spatial gaze. An immersive experience requires very high-resolution video – well beyond HD. These requirements lead to the need for adaptive strategies which allow to transmit, switch between, and decode multiple high resolution image streams (each covering a certain spatial region) while decoding the video streams with imperceptible latency.

3.6 Image sensor compression (raw-Bayer)

Direct image sensor compression, for example in still image and movie RAW cameras, has several advantages compared to the usual compression in the RGB domain.
First, in terms of complexity, less data needs to be processed; i.e. in a Bayer pattern coming from a sensor, there is 3 times less data to process than the subsequent RGB image obtained after debayering. This induces a drastic reduction in terms of buffering requirements. Moreover, performing the compression as close as possible from the sensor allows to bring all advantages of the JPEG XS compression to segments of dataflows that were still using uncompressed data so far. This makes JPEG XS beneficial for even more use cases and markets, as listed below.
In all use cases, capturing and processing of high dynamic range (HDR) images becomes important to take full advantage of the high sensitivity of the image sensors. Moreover, capturing scenes in HDR allows for advanced post-processing and image analysis workflows.
The main target applications for image sensor compression are described below.

3.6.1 Broadcasting and high-end cameras

The camera control unit (CCU) is an essential unit in a live television broadcast chain. It is responsible for powering the professional video camera, handling signals sent over the camera cable to and from the camera, and can be used to control various camera parameters remotely. CCU usually converts the RAW data camera stream in YUV 4:2:2 signals, by managing black level adjustment, denoiser, color conversion, white balance, debayering, gamma curve and chroma-subsampling. The video link between the camera and the CCU is currently always uncompressed RAW data, with a minimal latency (<<1 ms). With the increase of resolution, the need of higher speed cameras and long distance remote CCU, there is a need of extremely low latency & low buffering raw compression with a compression ratio from 2:1 (supporting even mathematically lossless on some images) to 4:1 (visual transparency). In addition, HDR cameras with a non-linear OETF are entering the market, which may require extra processing steps to allow efficient compression.

3.6.2 Prosumer and consumer cameras (including mobiles)

With the increase of the sensor resolution, the high-volume market faces the challenge to always offer better high-quality products while keeping a low cost and low power approach. Decreasing information transfers by compressing the image-sensor data is an attractive solution to reduce power consumption. To be successful in this competitive environment, we must provide the smallest sensor-data image compression scheme, with the lowest memory as possible, to reduce the fabrication cost, and more importantly the power consumption and heat dissipation. In the sectors where the uncompressed image is the norm, the image quality is very sensitive, and the compression must be transparent to be an appealing option. The required memory size of a combined encoder-decoder would usually be 8 lines of samples of an uncompressed raw Bayer image. Finally, cameras using non-Bayer type of sensors (e.g. non RGGB or using other filter types and subpixel layouts) have entered the market.

3.6.3 Machine vision

Machine Vision Systems typically have cameras for capturing images and powerful image processing systems for image analysis. The cameras themselves are separated from the image processing systems due to mounting and dimension constraints. It is beneficial to use the cameras only as sensor devices with as few computation capabilities as possible. A debayering in the camera would increase the computational load and power of such devices. Only the compression should be added. For automatic analysis a debayering to RGB images is not necessary. In terms of latency, ultra-low latency is not necessarily a requirement, but it has to be a fixed and constant value.

3.6.4 Automotive industry

One of the most important applications in the automotive industry is autonomous driving. For this application multiple sensors are combined, and their data transmitted to a so-called ECU (Electronic

central unit) in which the data from the sensors are jointly analysed and related actions calculated. Data from these sensors need to be processed with a maximum responsiveness, therefore implying a very low latency along the whole dataflow. In terms of implementation and given the number of sensors, power consumption needs to be constrained as much as possible because of thermal considerations and the necessary operation in all kinds of climatic conditions. The cabling inside the car is critical, e.g. fiber is not likely to be used as they are not as robust as copper cable and cannot be bent. High resolution cameras and limited interface bandwidths require compression to reduce the bandwidth during live transmission and/or aggregating multiple sensor streams on one port/cable. For automatic analysis a debayering to RGB images is not important. Given the infrastructure encountered in a car, multi-platform implementations of such compression schemes would be required, including GPU, FPGA, and ASIC.
In addition, the automotive industry often uses special types of CFA sensors that do not have typical RGGB-style of color filters, but rely on more specialized sensor layouts and color filters (e.g. RYYB). With such sensors, the sensor readout is not necessarily linear to the captured light intensity.
The above-described use cases can use different types of interfaces, each with its own throughput. The table hereunder indicates the targeted interfaces and their corresponding application areas.

Table 2 — Examples of access rate requirements for typical physical storage devices

Interfaces	Nominal throughput (Gbps)	Application areas
MIPI A-PHY	2, 4, 8, 12, 16	• Vehicles
MIPI C/D-PHY	~6/lane	• Vehicles • Mobiles • Sensors • Pro-AV
Ethernet Cat-5e/Cat-6e	up to 5 for Cat 5e up to 10 for Cat 6e	• Vehicles • Broadcast • Pro-AV • Machine vision
DDR3/DDR4 access	~23	• Pro-AV • Mobiles
PCI Express	~16-32/lane	• Pro-AV • Mobiles
SATA	~3-6	• Pro-AV • Mobiles
LVDS	~1-3	• Broadcast • Vehicles • Machine vision

3.7 Transport over 5G networks

5G network technologies bring a variety of new opportunities for the media sector, such as the incorporation of UHD video and immersive AR/VR content into streaming applications to provide a reliable and high-quality user experience. Ultra low end-to-end latency is a crucial requirement to support real-time interactivity and provide a seamless immersive user experience. The following table provides the targeted bandwidths and throughputs for the transport of video content over 5G (3GPP) with XS.

Table 1 — Examples of bandwidth requirements and compression ratios for various commonly used video formats

Video stream	Video throughput [1]	Physical link	Available throughput	Comp. ratio
4K@30p 420 8-bit	2.95 Gbps	5G (3GGP)	1 Gbps	~3
4K@30p 420 8-bit	2.95 Gbps	5G (3GGP)	0.5 Gbps	~6
4K@30p 420 8-bit	2.95 Gbps	5G (3GGP)	0.25 Gbps	~12
4K@60p 420 8-bit	5.92 Gbps	5G (3GGP)	1 Gbps	~6
4K@60p 420 8-bit	5.92 Gbps	5G (3GGP)	0.5 Gbps	~12
8K@30p 420 8-bit	11.94 Gbps	5G (3GGP)	1 Gbps	~12
[1] Gbps = Gigabit per second				

3.7.1 Video over 5G networks

The high bandwidth of 5G networks allows for UHD video streaming over mobile networks. However, real-time applications are usually limited because of the codec complexity. A codec with low complexity and lightweight compression capability will enable real-time transmission and reproduction on mobile terminals using 5G networks.

3.7.2 Interactive and immersive applications over 5G networks

Interactive and immersive applications require ultra-low latency and efficient video distribution for an enhanced quality of experience. Hence, they demand lightweight compression of the source content in response to the different interactions, and very high resolution for improved immersion using HMDs and mobile devices.

4 Requirements

This section presents the overall requirements that should be met by the proposals so as to be suited for the use cases described above. These requirements are split between “core coding requirements” and “extended features”.

4.1 Core coding requirements

4.1.1 Uncompressed domain (image characteristics)

Image resolution	from VGA (640x480) up to 10K
Component subsampling	4:0:0, 4:2:0, 4:2:2, 4:4:4, 4:2:2:4, 4:4:4:4
Component type	grayscale, RGB, YCbCr, CFA (e.g. Bayer patterns) ● Input type of the encoder shall match the output type of the decoder. ● Internal color space conversion is permitted.
Component bit-depth	8 to 16 bits per component (bpc) integer

Frame rate	from 24 fps to 120 fps, progressive content
Content	natural, synthetic, screen content
Color spaces support	e.g. Rec. BT.709 [1], Rec. BT.2020 [2], P3D65 [3], LogC, etc.
Transport curves support	e.g. Standard Gamma, Hybrid Log Gamma (HLG), and PQ Gamma.
Full HDR support	see color spaces, transport curves and bit-depth

4.1.2 Compressed domain (bitstream properties)

Image quality *(see section 2.4)*	● Image quality between high and near-visually lossless with compression ratios of 15:1 to 6:1. ● Image picture quality between near-visually lossless and transparent at compression ratios of 15:1 to 2:1. ● Flickering-resilient visually lossless quality at 3 bpp. ● Mathematically lossless visual quality at an approximate compression ratio of 2:1.
Bit rate allocation *(see section 2.3)*	● Variable bit rate (VBR) and constant bit rate (CBR) support. ● The ability to define a strict compressed size per frame (including overhead and signaling). ● Guaranteed avoidance of exceeding the target rate.
Self-contained compressed frame	A compressed frame shall contain all information required to completely recover the corresponding uncompressed frame (resulting in a viewable image).
HDR support	Signaling of image parameter values regarding HDR and color spaces.
Multiple generation robustness	Robustness to at least 7 encoding-decoding cycles (no quality degradation), while allowing intermediate processing operations (overlay, crop, editing, pan and scan).

4.1.3 Design and implementation

- **Ultra-low latency** (see section 2.2):
  - o A maximum algorithmic latency of 32 video lines for a combined encoder-decoder suite. In the context of CFA data processing, a “line” must be interpreted as a line of “super-pixels” as defined in ISO/IEC 21122-1 2nd Edition.
  - o The codestream syntax shall be defined in such a way to not prevent an encoder and decoder implementation to fit in a targeted FPGA device for an end-to-end latency of 32 lines.
- **Low complexity** (see section 2.1):
  - o The algorithm shall be defined in such a way to allow for low complexity implementations on multiple platforms (in software and hardware).
  - o For hardware: As an indication, to process real-time 4K 4:4:4 8bit 60p with a compression ratio compliant with the requirements, either encoder or decoder should require less than 50% of an FPGA similar to Xilinx Artix 7 [5] or 25% of an FPGA similar to Intel/Altera Cyclone 5 [6].

   - For software: An optimized software implementation should be able to real-time process 4K 4:4:4 8bit 60p on an Intel i7 (of 2016) processor, or equivalent.
- **End-to-end parallelization** (see section 2.1):
   - Support CPU (threads/SIMD), GPU, FPGA and ASIC parallelization to avoid any serial bottlenecks in the encoding and decoding process.
- **Implementation scalability**:
   - The resources required by the encoder and the decoder shall scale depending on required throughput.
- **No external memory dependency**:
   - Support hardware implementations without external memory.
- **Multiple platform interoperability** (see section 2.1):
   - Implementations of different types shall be allowed/capable to produce the same codestream and be interoperable.
- **Deterministic encoding**: The codestream syntax shall be defined to allow an encoder to deterministically generate a codestream that only depends on the currently encoded frame, while still meeting latency and complexity requirements.
- **Configurability**:
   - Image size, frame rate, bit-depth (bpc), component type, subsampling.
   - Targeted compressed bit rate (bpp), VBR or CBR, maximum compressed size per frame.
   - Capability to enable/disable optional features (see section 4.2), as disabled features might facilitate smaller hardware or software footprints.

4.2 Extended features

- **Extended sensor format support**:
   - Alternative CFA patterns
      - RGBW, RGBE, RCCB, RYYB, RGG'B, RCCC, etc.
      - 2x2 patterns, 4x4 patterns, rotated patterns, etc.
   - Special sensor support:
      - Multiple (i.e. 2 or 3 per component) sensitivity cells per pixel, combined into one output.
      - Manage transfer curve for pre-emphasis phase (OETF).
      - Might require 20-bit support at the input.
- **HDR support**:
   - Support for ISO/IEC 60559 16-bit floating point [7] component data.
- **Efficient spatial (random) access** (targeted at file-based workflows):
   - No full decode required.
   - Avoid impacting any low-complexity properties outside of use cases that do not need this.
   - Coarse grained:
      - To slices in the codestream (technically it refers to the SLH marker that misses a length marker, which was done on purpose to allow for low latency).
      - To columns within the slices.
   - Fine grained:
      - To localized spatial regions (smaller than slices or columns).
   - Partial updating of codestreams (local editing operations).
- **Improved compression performance**: Reduce the bit rate by 50%, while keeping the same quality as with XS high profiles. This should be feasible by means of the following relaxed constraints:
   - Allow frame quality degradation on scene cuts, with the lowest frame quality bound to the quality achievable with XS high profiles at the same bit rate.
   - Relaxed low latency (see section 2): up to maximally 1/4th of a frame end-to-end.

   - Allow a very low end-to-end latency (below 10 ms) implementation suitable for interactive applications over 5G networks (see Table 3).
   - Allow image qualities between high and near-visually lossless, as defined in JPEG AIC-3 for applications over 5G networks (see Table 3).
   - Relaxed low complexity (see section 2.1):
     - For FPGA, allow additional computational complexity of up to 100% on top of core requirements (see section 4.1.3). For software this means that the execution speed shall be at most 50% slower for highly parallelized implementations.
     - Minimize the algorithmic memory requirements to allow using on-chip memory as much as possible (e.g. block RAM on FPGA, cache on CPU, local memory on GPU). The requirements for on-chip memory shall not exceed 100% of the core memory requirements.
   - Allow usage of external memory, provided that memory bandwidth is under control (e.g. using compressed frame buffer at 12 bpp).
     - For example: For 4K60, a memory configuration of DDR3/DDR4 running at 800 MHz with a width of 16-bit and a max throughput of uncompressed content (12 bpp).
     - Allow a framebuffer.
       - The algorithm shall provide a way to put an upper boundary on the number of frames it will take to ensure that each pixel value is updated independently of the frame buffer. This is required to enable timely and full (or lower quality) reconstruction of the frames, e.g. for when arbitrary switching into an ongoing connection, or to limit error propagation. One example would be a 200 ms maximum switching time for full reconstruction (not to be confused with latency).
       - The algorithm shall provide an optional mechanism to update active regions in the frames more frequently than less active regions in the frame.
       - The algorithm shall have the means to constrain the memory bandwidth between the codec and the frame buffer memory. A scalable solution in terms of the selected memory bandwidth is preferred.
       - The algorithm shall be scalable in terms of the size of the frame buffer.
   - Codestreams do not have to be self-contained anymore to allow for inter-frame correlation, while keeping the complexity of encoder and decoder similar.
   - The improved compression performance decoder must be capable of decoding all codestreams up to the high profiles. Existing coding tools of the high profiles should be supported as a subset.
   - Multi-generation robustness is not a mandatory requirement in this scenario.
- **Transcoding friendly support**: Facilitate support for efficient transcoding to/from XS (not for live).
   - Support for other wavelet kernels (e.g. 9/7).
   - ICT color transformation handling.
   - Allow more than 2 vertical decomposition levels.
   - VBR allocation (no loss of quality).
   - Present as an XS profile (allowing certain complexity) that provides high performance encoding and decoding.
- **Miscellaneous features**:
   - Robustness to post-processing operations (such as subsequent editing operations, color transform or gamma conversion that shall not induce visual artefacts). Especially in the case of HDR content, the compression workflow should include processing steps to ensure that

the image quality remains ideal, even when HDR transform curves are applied for rendering of the reconstructed images.

- Error robustness: independently from error protection mechanisms available at transport level, the codestream should facilitate implementations to minimize the impact of bit errors and packet losses.
- Handling of different electro-optical transfer functions (EOTF): the proposed algorithm should optimize its performance by considering the transfer function being used by the content to be processed.
- Image proxy support: Support for extracting multi resolutions from the generated codestream.
- Support handling of metadata signaling for information like color gamut, white point, black level, toe threshold or transfer curves.

Bibliography

[1] ITU-R BT.709-5 (2002), “Parameter values for the HDTV standards for production and international programme exchange”. https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.709-5-200204-I!!PDF-E.pdf.

[2] ITU-R BT.2020 (2012), “Parameter Values for ultra-high definition television systems for production and internal programme exchange”. http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.2020-0-201208-I!!PDF-E.pdf

[3] SMPTE ST 428-1:2006 “D-Cinema Distribution Master – Image Characteristics”. P3 with D65 white point.

[4] ISO/IEC 29170-2, “Information technology – Advanced image coding and evaluation methodologies – Part 2: Evaluation Procedure for nearly lossless coding”.

[5] Xilinx Artix 7, http://www.xilinx.com/products/silicon-devices/fpga/artix-7.html

[6] Altera Cyclone V, https://www.altera.com/products/fpga/cyclone-series/cyclone-v/features.html

[7] ISO/IEC 60559:2020, “Information technology – Microprocessor Systems – Floating-Point arithmetic”.

[8] OpenEXR, https://www.openexr.com

[9] M. Testolina, E. Upenik, J. Sneyers, and T. Ebrahimi, ‘Towards JPEG AIC part 3: visual quality assessment of high to visually-lossless image coding’, in Applications of Digital Image Processing XLV, 2022, vol. 12226, pp. 90–98.

