Real Time Image Processing Journal manuscript No.
(will be inserted by the editor)

Alessandro Artusi · Rafał K. Mantiuk · Thomas Richter · Philippe Hanhart · Pavel Korshunov · Massimiliano Agostinelli · Arkady ten · Touradj Ebrahimi

Overview and Evaluation of the JPEG XT HDR Image Compression Standard

Received: date / Revised: date

Abstract Standards play an important role in providing a common set of specifications and allowing interoperability between devices and systems. Until recently, no standard for High Dynamic Range (HDR) image coding had been adopted by the market, and HDR imaging relies on proprietary and vendor specific formats which are unsuitable for storage or exchange of such images. To resolve this situation, the JPEG Committee is developing a new coding standard called JPEG XT that is backwards compatible to the popular JPEG compression, allowing it to be implemented using standard 8-bit JPEG coding hardware or software. In this paper, we present design principles and technical details of JPEG XT. It is based on a two-layers design, a base layer containing a Low Dynamic Range (LDR) image accessible to legacy implementations, and an extension layer providing the full dynamic range. The paper introduces three of currently defined profiles in JPEG XT, each constraining the common decoder architecture to a subset of allowable configurations. We assess the coding efficiency of each profile extensively through subjective assessments, using 24 naïve subjects to evaluate 20 images, and objective evaluations, using 106 images with five different tone-mapping operators and at 100 different bit rates. The objective results (based on benchmarking with subjective scores) demonstrate that JPEG XT can encode HDR images at bit rates varying from 1.1 to 1.9 bit/pixel for estimated mean opinion score (MOS) values above 4.5 out of 5, which is considered as fully transparent in many applications. This corresponds to 23-times bitstream reduction compared to lossless OpenEXR PIZ compression.

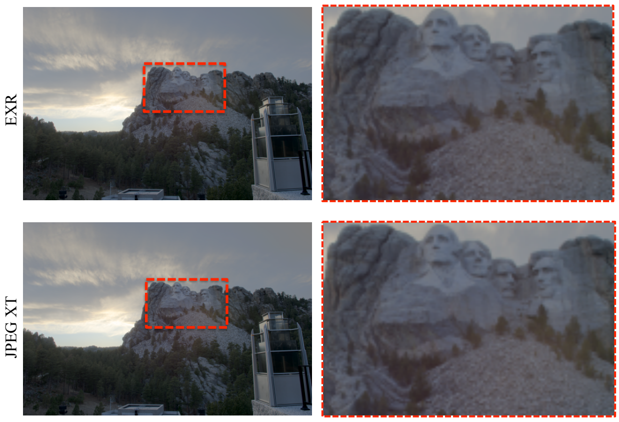

Fig. 1 Original (EXR) and compressed and decoded image (JPEG XT), both tonemapped and scaled for the purpose of printing. The original image size in the OpenEXR (PIZ compression) format is 39,4 MB (24.0 bpp) large, the compressed size is 3,4 MB (2.0 bpp), less than one tenth of the original file size.

1 Introduction

Despite a rapid increase of scientific activities and interests in High Dynamic Range (HDR) imaging, its adoption by industry is rather limited. One of the reasons is the lack of a widely accepted standard for HDR image coding that can be seamlessly integrated into existing products and applications. While standard formats such as JPEG 2000 and JPEG XR offer support for HDR image representations, their adoption requires a certain investment not always affordable in existing imaging ecosystems, and more difficult transitions, as they are not backward compatible with the widely popu-

A. Artusi
University of Girona
Edifici P4, 17071 Girona (Spain)
E-mail: artusialessandro4@gmail.com

R. K. Mantiuk
University of Cambridge, Computer Laboratory
15 JJ Thomson Avenue, Cambridge, CB3 0FD, UK
E-mail: mantiuk@gmail.com

Th. Richter
University of Stuttgart
Computing Centre
Allmandring 30A
E-mail: richter@tik.uni-stuttgart.de

lar JPEG image format (Pennebaker and Mitchell 1992; Wallace 1992). Instead, most digital camera and mobile phone manufactures offer an “HDR mode”, which is based on a vendor-specific proprietary technology. This situation creates a “vendor lock-in” problem for consumers, making it difficult to efficiently use images produced by such cameras in practice. While cameras typically offer an option to generate a tone-mapped 8-bit JPEG version from the capture HDR image, it cannot be considered as an original HDR “digital negative” and, hence, is not optimal for editing and creative enhancements.

To resolve this problem, in 2012, the JPEG Committee formally known as ISO/IEC JTC1/SC29/WG1, issued a “call for proposals” to which 6 organizations responded, namely, Dolby, EPFL, University of Stuttgart, Trellis Management, VUB and University of Warwick. As a result, JPEG XT was initiated as a new work item and a first set of requirements for its potential applications was identified. An important requirement was the possibility for any legacy JPEG decoder to be able to recover a Low Dynamic Range (LDR) version of the coded HDR image, resulting in a two-layer design of a base LDR and an extension codestream. Another important requirement was to impose both base and extension codestreams to use legacy JPEG compression tools in order to facilitate implementations. Compression efficiency was also considered as a third objective.

JPEG XT standard defines a common codestream syntax and a common decoder architecture. To make practical implementations easier, the set of the coding tools offered by the standard can be restricted to smaller subsets denoted as *Profiles*. This paper focuses on three JPEG XT profiles, referred to as profiles *A*, *B*, and *C*. Each profile offers a technical solution for coding HDR images considering additional requirements for different applications.

In this paper, we present the overall design architecture of JPEG XT and discuss the above mentioned three profiles. We also extensively evaluate the performance of JPEG XT. Subjective assessment of JPEG XT profiles is made using 24 naïve subjects evaluating 20 different HDR images coded at 4 different bit rates and displayed on a SIM2 HDR47E S 4K monitor. Based on the results of the subjective evaluations, twelve different quality metrics, including several variants of PSNR, RMSE, and SSIM, as well as, SNR and HDR-VDP-2 (version 2.2) have been benchmarked. The benchmarking showed that HDR-VDP-2 offers the highest correlation with subjective results. Therefore, HDR-VDP-2 was applied to a total of 106 different images from several publicly available image datasets that were coded at 100 different bit rates. In this evaluation, we also assessed the influence of five different commonly used tone-mapping operators, including simple *gamma* operator, *drago03* (Drago et al 2003), *reinhard02* (Reinhard et al 2002), *mai11* (Mai et al 2011), and *mantiuk06* (Mantiuk et al 2006b), on the performance of different profiles of JPEG XT.

In summary, the following are the main contributions of this paper: The first such extensive and detailed description of JPEG XT standard and its three profiles (Section 3); a comprehensive subjective assessment of three JPEG XT profiles using 20 HDR images (Section 6); benchmarking and statistical analysis of 12 objective metrics for HDR images using the results of subjective assessment (Section 6.2); a large-scale objective evaluation and performance analysis of JPEG XT using HDR-VDP-2 on 106 images (Section 7). Finally, real time issues, related to the JPEG XT, are discussed.

2 Related Work

Compression of high bit-depth still images, such as HDR images, has been investigated in the past. JPEG 2000 and JPEG XR have been proposed to overcome the limited bit-depth of the dominant standard for photographic images, the legacy coding system ISO/IEC 10918/T.81 widely known as JPEG format. Both JPEG 2000 and JPEG XR standards can represent HDR images when used in combination with an appropriate pixel encoding, such as logLuv (Ward-Larson 1998; Pattanaik and Hughes 2005) or perceptual quantization (Mantiuk et al 2004; Miller et al 2013). However, those standards have not been adopted by the digital photography market. As JPEG is currently *de facto* the most popular imaging format, it is believed that an HDR image coding format should be backward compatible with the legacy JPEG format to facilitate its adoption and inclusion in current imaging ecosystems.

First attempts to design a coding system for HDR still images that would also provide backward compatibility were (Spaulding et al 2003) and (Ward and Simmons 2006). The latter, known as JPEG-HDR, also proposed a software implementation which made it popular for compression of HDR images among some HDR enthusiasts. Minor limitations of that format were the lack of support for Wide Color Gamut (WCG) and lack of lossless coding. Other JPEG backward-compatible compression schemes have also been proposed for HDR video (Mantiuk et al 2006a) and for HDR images (Chen et al 2006; Korshunov and Ebrahimi 2013).

Recognizing the lack of a standard for compression of HDR images that is backward compatible with JPEG format, the JPEG Committee started a new work ISO/IEC 18477 also known as JPEG XT. Prototypes implementing JPEG XT architecture have been described and analyzed in literature as early as in 2013 (Richter 2013a,b). This initiative has attracted interest from researchers in academia and industry, leading to several studies assessing its performance.

The work by Pinheiro *et al.* (Pinheiro et al 2014) compared four tone-mapping operators in how they affect performance of three profiles of JPEG XT, when used to generate the base layer of a compressed image. This evaluation demonstrates the sensitivity of the compression results to the choice of the tone-mapping operator in the base layer and showed that profiles perform consistently at different bit rates when SNR and FSIM metrics were used for measurements. Other studies were mostly limited to the performance evaluation of only one of the three available profiles in JPEG XT (Mantel et al 2014; Hanhart et al 2014a). The work by (Mantel et al 2014) presented a subjective and objective evaluation for profile C. The objective grades were compared to subjective scores concluding that the MRSE metric provides best prediction performance. The authors of (Hanhart et al 2014a) investigated the correlation between thirteen well known full-reference metrics and perceived quality of compressed HDR content. Their evaluation was performed only on profile A of JPEG XT, cf. Section 3. In contrast to (Mantel et al 2014) their results showed that commonly used metrics, e.g., PSNR, SSIM, and MS-SSIM are unreliable in prediction of perceived quality of HDR content. They concluded that two metrics, HDR-VDP-2 and FSIM, predicted the human perception of visual quality reasonably well. The main limitation of these two studies is in the small number of images used in their experiments, which was limited to six and five, respectively. The study by (Valenzise et al 2014) compared the performance of three objective metrics i.e. HDR-VDP, PSNR and SSIM, when considering HDR images compressed with JPEG XT. The results of this study showed that simpler metrics can be effectively employed to assess image fidelity for applications such as HDR image compression.

Crowdsourcing has also been used to evaluate performance of JPEG XT coding standard (Hanhart et al 2014b). In particular, the feasibility of using LDR versions of original HDR content obtained with tone-mapping operators was investigated. This evaluation showed that some tone-mapping operators are more suitable for evaluation of HDR image compression.

Lately (Korshunov et al 2015) have provided a publicly available dataset of 20 HDR images and corresponding versions compressed at four different bit rates with three profiles of the upcoming JPEG XT standard for HDR image compression. The images cover different scenes, dynamic ranges, and acquisition methods. The dataset also includes Mean Opinion Scores (MOS) for each compressed version of the images obtained from extensive subjective experiments using SIM2 HDR monitor.

In this paper, we present an extensive series of objective and subjective evaluations for profiles A, B and C of JPEG XT. In contrast to previous work, we have used a large data set of challenging HDR content, resulting in a total of 106 images for objective evaluations, of which a subset of 20 images were deployed for subjective assessments. Twelve objective metrics are compared with the results of the subjective study by a correlation test, and five different tone-mapping operators have been run to generate LDR content as base layer for the JPEG XT encoders. To our best knowledge, this work is the most complete analysis of JPEG XT performance ever produced.

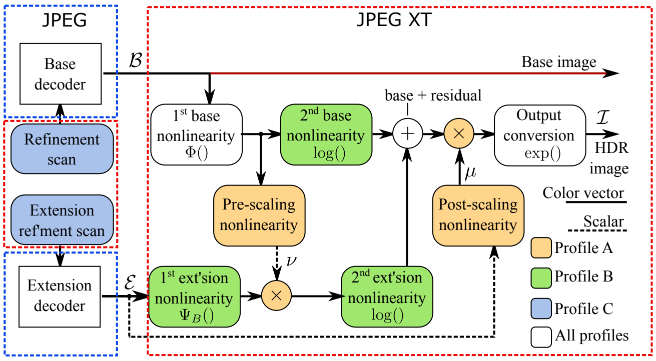

Fig. 2 The simplified JPEG XT standard decoder architecture: (blue-dashed line) is the legacy coding system ISO/IEC 10918/T.81 known as JPEG format - (red-dashed line) are the additional components that define the new JPEG XT standard.

3 The JPEG XT Standard

The JPEG XT standard currently consists of eight parts, with a ninth under preparation: Part 1 (Husak W., Richter T. to appear) defines the core coding technology, which is the legacy JPEG specifications as it is used and known today. Part 3 (Richter T., Schelkens P., Ishikawa T. to appear) defines an extensible and flexible container format extending legacy JPEG and the ISO-based media format. Part 6 (Richter T., Ogawa S. to appeara) specifies technology for coding of integer sample formats between 8 and 16 bits precision. Part 7 (Richter T., Artusi A., Agostinelli M. to appear) covers coding of images in a *HDR representation*, e.g. dynamic range requiring floating point samples, as discussed in this paper. Part 8 (Richter T., Ogawa S. to appearb) finally combines coding technologies from Parts 6 and 7 to allow lossless coding of intermediate and HDR image representations. Parts 4 (Richter T., Ten A., Artusi A. to appeara) and 5 (Richter T., Ten A., Artusi A. to appearb) will define conformance testing and provide a reference implementation, whereas Part 2 (Richter T., Husak W., Ninan A., Ten A., Jia W., Korshunov P., Ebrahimi T., Artusi A., Agostinelli M. to appear) supplies a legacy syntax for a subset of the tools specified in Part 7.

The functional blocks of standard decoders from parts 3 to 8 can be merged into one consistent diagram describing the overall decoder functionality, depicted in Figure 2. Roughly, a JPEG XT image consists of a legacy codestream using the 8 bit Huffman coding mode of ISO/IEC 10918-1, and an extension (residual) codestream inserted in application markers extending the precision of the image to a target precision. The box in the top-row of Figure 2 decodes the legacy codestream to form the base LDR image; the extension decoder in the bottom row works likewise and uses either an 8 bit or a 12 bit mode of the traditional JPEG standard. Both images are merged together by an addition and/or multiplication, combined with further processing by the functional blocks in the middle and right sides of the diagram. These are based only on two elementary types of operations: A non-linear transformation, acting independently on each image component, and a linear 3×3 transformation implementing either color transformations or inverse decorrelation transformations. Both operate on a pixel-per-pixel basis, and no other inverse transforms beyond DCTs in the base and the extension decoding paths are used. Hence, a JPEG XT decoder can be simply implemented by employing two standard JPEG decoders and one additional processing merging their outputs pixel by pixel.

3.1 Profiles

Three profiles of Part 7 for HDR floating point coding are described in more details hereafter. These are also the profiles whose coding performance is discussed in the rest of the paper. As in many standards, profiles constrain the choices of coding parameters and functional blocks allowed in a codestream conforming to such profiles; while implementation of a JPEG XT covering requirements of all potential applications are possible, decoders conforming to a specific profile are only required to implement a subset of the functional blocks. Currently, Part 7 defines four profiles A, B, C and D, of which profile D is a very simple entry-level decoder that allows a 12 bit mode compatible to the 8 bit Huffman mode of JPEG while offering a precision similar to the 12 bit mode of legacy JPEG. It is not considered in this paper.

What is common to all JPEG XT profiles is that they all take into account the nonlinearity of the human visual system (HVS). While the relation between the physical luminance (stimulus) and response can be roughly approximated by a power function in the LDR domain, known as “gamma correction”, this relation can be more appropriately modeled by a logarithmic function for HDR. The approximately logarithmic behavior between stimulus and response is also known as the Weber-Fechner law. JPEG XT makes use of this relation by representing an HDR image as a sum of base and extension images in the logarithmic domain, or their product (ratio) in the linear domain[1].

In all profiles, the *Base Image* $\mathcal{B}$ is always represented in regular JPEG codestream that decodes to a low-dynamic range, 8 bits per sample image in the ITU BT.601 RGB colorspace if the JPEG XT specific side information is ignored. Especially, decoding $\mathcal{B}$ includes the inverse decorrelation transformation from YCbCr to RGB that at encoding level takes into account for the decorrelation of the three RGB color channels. This is equivalent to equation 3 for both base and extension layers. The *Extension Image* $\mathcal{E}$ includes additional information to reconstruct from $\mathcal{B}$ and $\mathcal{E}$ a high-dynamic range image $\mathcal{I}$. While $\mathcal{E}$ is also decoded by the JPEG algorithm, the transformation from YCbCr to RGB is here *not* part of the decoding and made explicit in the formulae that follow. $\mathcal{E}_0$ denotes the (scalar) luma component of the extension image and $\mathcal{E}^\perp$ the extension image projected onto the chroma-subspace, i.e. $\mathcal{E}$ with its luma component set to zero. Which coding tools of the overall JPEG XT infrastructure are used to merge $\mathcal{B}$ and $\mathcal{E}$ together is then profile dependent.

In profile A, the HDR image $\mathcal{I}$ is represented as a product of a luminance scale μ and the base image $\mathcal{B}$ after inverse gamma correction[2] through Φ_A (*1^st base nonlinearity* in Figure 2). μ is a scalar function of the luma component of the extension image in the *Post-scaling nonlinearity* block. Formally, the reconstruction algorithm for profile A then reads as follows:

$$\begin{aligned}\mathcal{I}(x,y) = \mu\big(\mathcal{E}_0(x,y)\big)\cdot\Big[C\,\Phi_A\big(\mathcal{B}(x,y)\big)\\ + \nu\Big(S\,C\,\Phi_A\big(\mathcal{B}(x,y)\big)\Big)\cdot R\,\mathcal{E}^\perp(x,y)\Big]\end{aligned} \quad (1)$$

where C and R are 3×3 matrices implementing color transformations. The matrix C transforms from ITU-R BT.601 to the target colorspace in the base image. If the target color space is, for example, ITU-R BT.2020, this matrix can be computed as

$$C = \begin{bmatrix} 1.544 & -0.320 & -0.228 \\ -0.567 & 1.375 & 0.013 \\ 0.023 & -0.055 & 1.214 \end{bmatrix} \begin{bmatrix} 0.640 & 0.290 & 0.150 \\ 0.330 & 0.600 & 0.060 \\ 0.030 & 0.110 & 0.790 \end{bmatrix} \quad (2)$$

where the first matrix is the conversion from XYZ to linear BT.2020, and the second matrix is the conversion from linearized BT.601 to XYZ.

R is an inverse decorrelation transformation from YCbCr to RGB in the extension image. Typically, it is identical to the conversion from YCbCr to RGB as defined by BT.601:

$$R = \begin{bmatrix} 1.000 & 0.000 & 1.402 \\ 1.000 & -0.344 & -0.714 \\ 1.000 & 1.772 & 0.000 \end{bmatrix} \quad (3)$$

S is a row-vector transforming color into luminance, and ν is a scalar function of this luminance value. Typically, $\nu(x) = x + \epsilon$ where ϵ is a “noise floor” that avoids an instability in the encoder for very dark image regions. Inverting the decoder equation (1) as necessary for encoding includes a division by $\nu(S\mathcal{B})$. Dark image areas

[1] Recall that $a \cdot b = \exp\big(\log(a) + \log(b)\big)$.

[2] Sample values are proportional to physical intensities after inverse gamma correction.

hence result in denominators to come close to zero. Even though the numerator is in such clases also close to zero, the computation of the quotient is then numerically unstable. Adding the noise floor ϵ prevents this problem.

Profile B follows a different strategy by splitting the image along the luminance axis into "overexposed" areas and LDR areas. The overall image $\mathcal{I}$ is then, in general, represented as the RGB component-wise quotient $\mathcal{B}/\mathcal{E}$ where $\mathcal{E}$ is the unity in areas that are captured in the LDR base image. In overexposed areas $\mathcal{E}$ values fall below 1 while $\mathcal{B}$ remains at its maximum value. Formally, the reconstruction algorithm is expressed as follows:

$$\begin{aligned}\mathcal{I}(x,y)_i &= \sigma \exp\Bigg(\log\Big(\big[C\,\Phi_B\big(\mathcal{B}(x,y)\big)\big]_i\Big)\\ &\qquad - \log\Big(\Psi_B\big(\big[R\,\mathcal{E}(x,y)\big]_i\big)+\epsilon\Big)\Bigg)\\ &= \sigma\frac{\big[C\,\Phi_B\big(\mathcal{B}(x,y)\big)\big]_i}{\Psi_B\big(\big[R\,\mathcal{E}(x,y)\big]_i\big)+\epsilon} \qquad (i=0,1,2)\end{aligned} \tag{4}$$

where i is the index of one of the RGB color channels, Φ_B is the inverse gamma correction, and Ψ_B a content-dependent nonlinearity applied in the extension. The decoder implements the subtraction from the first two lines of the above equation in the summation block in Figure 2. The exponential and logarithmic functions are realised by the *2^{nd} base*, *2^{nd} extension nonlinearities* and *output conversion* blocks in Figure 2. Φ_B and Ψ_B are base and extension nonlinearities. The additional scale σ can be understood as an exposure parameter that scales the luminance of the output image to optimize the split between base and extension images.

Profile C also employs a sum to merge base and extension images, but here Φ_C not only approximates an inverse gamma transformation, but implements a global inverse tone-mapping procedure that approximates the (possibly local) tone mapping operator (TMO) that was used to create the LDR image, similar to (Mantiuk et al 2006a). The extension is encoded in the logarithmic domain directly, avoiding an additional transformation. Finally, log and exp are substituted by piecewise linear approximations that are implicitly defined by re-interpreting the bit-pattern of the half-logarithmic IEEE representation of floating-point numbers as integers. It is then easily seen that this simple "casting" between number formats implements two functions $\psi\log$ and $\psi\exp$ that behave approximately like their precise mathematical counterparts, though they provide the additional advantage of being exactly invertible (Richter 2014). The reconstruction algorithm for profile C then reads:

$$\mathcal{I}(x,y) = \psi\exp\Big(\hat{\Phi}_C\big(C\,\mathcal{B}(x,y)\big)+R\,\mathcal{E}(x,y)-2^{15}(1,1,1)^T\Big) \tag{5}$$

where $\hat{\Phi}_C(x) = \psi\log\big(\Phi_C(x)\big)$, in which Φ_C is the global inverse tone-mapping approximation. 2^{15} is an offset shift to make the extension image symmetric around zero. The codestream never specifies Φ_C directly, but rather includes a representation of $\hat{\Phi}_C$ in the form of a lookup-table, allowing to skip the time-consuming computation of the logarithm; additional implementation details for all profiles relevant for real-time performance are discussed in section 4.

The interesting property of profile C is that it allows lossless coding as its decoding algorithm only requires invertible integer operations; the only change in the lossless mode is that the DCT in the reconstruction of $\mathcal{E}$ is bypassed, and the implementation of the base layer DCT is fully specified by the standard.

Profile C also adds the option of *refinement scans*, which increases the bit-precision in the DCT domain by adding least-significant bits by a method similar to the progressive mode in legacy JPEG. Similar to the extension layer, they are hidden from legacy applications. In the tests, they were only used in the extension layer.

3.2 Profiles Configuration

While each of the profiles is already a specialization of the general coding architecture of Figure 2, the standard still allows a lot of freedom within each of them. To make results comparable and to harmonize the profiles, we decided to select one common configuration for all tests in this paper: The base layer always uses 4:2:0 chroma-subsampling, as it is traditionally employed in JPEG compression. To allow optimal quality, we decided to enforce 4:4:4, i.e. no chroma-subsampling, for the extension layer. All implementations enabled optimized Huffman coding, i.e. used a two-pass encoding to identify the optimal Huffman alphabet. Nevertheless, small deviations in the base layer rate-distortion can be observed because implementations used differing legacy JPEG engines. Profile C in particular uses a 12 bit extension (8 bit legacy coding plus four refinement bits) for which no example Huffman table has been listed in the legacy JPEG; it should be noted, however, that the rate-distortion curve of the 8-bit and 12-bit extension mode lie exactly on each other as quantization loss dominates, except that the 12-bit mode allows profile C in particular to extend this curve towards higher bitrates and higher qualities, allowing scalable lossy to lossless coding.

Despite these choices, we imposed no further restrictions or requirements on the encoder, though requested all vendors to supply their recommendations for optimal coding performance. Like many JPEG and MPEG standards, JPEG XT itself does not specify the encoder either and only imposes the requirement that it creates a syntactically correct codestream that describes the desired image with suitable precision. Such error bounds will be defined in JPEG XT Part 4.

	Profile A	Profile B	Profile C
Additions	9	3	6
Multiplications	12	6	0
Look-up	0	0	3
Functions	4	6	0

Table 1 Number of additions/subtractions, multiplications/divisions, table look-up operations and functions. This does not include the operations necessary to transform from YCbCr to RGB, as this is formally not part of the JPEG standard. Scalar functions can be substituted by table look-up operations after scaling. The functions are typically the inverse gamma, exponential or logarithm, as required per profile for a typical implementation. The overall complexity of the encoder/decoder is $O(N)$.

4 Real-Time Issues

The JPEG core decoding algorithm, and hence JPEG XT consists of a Huffman decoder followed by dequantization and a discrete cosine transformation, see (Pennebaker and Mitchell 1992) Figure 2. JPEG XT stacks two conventional JPEG codecs and merges the outputs by post-processing; the (simplified) post-processing chain is depicted in Figure 2 where each profile selects a sub-set of the post-processing tools defined by the standard, see section 3. It is, however, important to note that all operations available as post-processing tools are static operations that do not depend on the pixel neighborhood or pixel position, and hence can be trivially parallelized.

Even more so, profile C of part 7 and part 8 (lossless) require only integer operations (Richter 2014) — in particular additions, subtractions and table lookup operations — and hence allow low complexity hardware implementations. Table 1 lists the operation counts for post-processing per pixel for some typical implementations of the algorithms introduced in section 3.

The overall complexity of the encoder and decoder is $O(N)$, where N is the number of pixels: The number of operations per block in the DCT domain is bounded by the Huffman en/decoder, the operations for the cosine transformation and the quantization. All these operations are required twice, once for the base and once for the extension layer, and hence the complexity for entropy coding and transformation is approximately doubled when compared to the legacy JPEG standard. The complexity for merging both layers into one single picture adds on top, but as the same operations are carried out for all pixels in parallel, another term of order $O(N)$.

JPEG XT part 7 does not offer an *online compression mode*, unlike conventional JPEG whose sequential mode allows encoding and decoding of data as they arrive (Pennebaker and Mitchell 1992; Wallace 1992). However as JPEG, JPEG XT allows the end-user to take concurrently to the storage process a series of HDR images with a DLSR camera, without the end-user being aware of it. The necessity to multiplex the base image and the extension image into one joint codestream also requires buffering of one complete image frame in coded or uncoded form, and hence inhibit online encoding: the complete codestream representing the extension layer needs to be known before it can be embedded into the baseline codestream. A similar constraint holds for decoding: A decoder needs at least to localize the first segment of the extension layer and the start of the first scan of the base layer to be able to decode the first 8×8 block in the HDR domain. This is due to a restriction of the codestream syntax of the legacy JPEG standard, which only allows for the inclusion of side channel data at scan boundaries. The side channel consists here on the encoded residual data in the extension layer and all the metadata to configure the decoder.

This meta data steer pre-processing on encoding that generates the extension layer from the (LDR,HDR) image pair, and also parametrize the corresponding post-processing at the decoder. For profile A, the parameters consist of the μ map, encoding the ratio between the LDR and HDR image. This map is typically a segment of an exponential function, but could also be represented by a look-up table. A profile B encoder needs to find an exposure value σ and a suitable Ψ_B map, i.e. a gamma value for the extension codestream, whereas profile C needs to estimate the inverse tone mapping $\hat{\Phi}_C$, which is then included as a look-up table.

As long as the relation between the HDR image and the LDR image is known, i.e. as long as the tone mapping that was used to generate the base image in first place is known, these parameters can be estimated without going over the input image pair. For example, the profile C $\hat{\Phi}_C$ map is the inverse of a global tone mapping operator combined with the pseudo-logarithm. In a typical application within a digital camera, this tone mapping operator is already part of the image processing chain as of today; it is necessary to create 8-bit/sample data suitable for the legacy JPEG encoder from the incoming sensor data whose bit-precision is usually higher.

In worst case, i.e. if the relation between HDR and LDR image is unknown to the encoder, the tonemapping, gamma values etc. need to be estimated by an additional scan over the source image (LDR,HDR) pair. Encoder parameters can then be found by an algorithm similar to that described in (Mantiuk et al 2006a) or (Ward and Simmons 2006). The overall encoder algorithm still remains $O(N)$, though the encoder complexity is approximately doubled as each pixel needs to be touched at least twice: Once to estimate the compression parameters, and once for the actual compression.

5 Test Conditions

The challenge of testing backward-compatible HDR compression is that the compression performance does not depend only on a single quality control parameter, but also on the quality settings for the base layer and on the choice of tone-mapping operator, which produces this layer. To fully understand the implications of those parameters, all their combinations were tested. We used the combination of 10 base quality levels × 10 extension layer quality levels × 5 tone-mapping operators × 3 profiles × 106 individual images, which results in a total of 159 000 conditions. However, such a large number of conditions clearly cannot be tested in a subjective experiment. Therefore, a subset of those conditions was used in a subjective experiment (Section 6) to find the most appropriate objective quality metric for a large-scale objective evaluation (Section 7).

Images For testing a set of 106 HDR images with resolutions varying from full HD (1920×1080) to larger than 4K (6032×4018) were selected. The dataset contained scenes with architecture, landscapes, portraits, frames extracted from HDR video, as well as computer generated images. All images were carefully selected from two publicly available datasets: Fairchild's HDR Photographic Survey[3] and HDR-eye[4] dataset of HDR images.

Then, the images were processed for subjective and objective evaluations as follows:

Set 1 (106 images for objective evaluation). HDR images, containing relative trichromatic values, cannot be directly used for objective evaluation. This is because the majority of objective HDR metrics are display-referred and expect that the values in images correspond to the absolute luminance emitted from an HDR or a LDR display, on which such images are displayed. Such metrics account for the fact that distortions are less visible in darker areas in an image. Because of that, this set was tone-mapped for a hypothetical HDR display of high contrast using a display-adaptive TMO (Mantiuk et al 2008). In order to introduce the minimum tone-mapping distortions and map to physically possible display, we assumed that the peak of such a hypothetical display is 4000 cd/m^2 and the black level is 0.02 cd/m^2.

Set 2 (20 images for subjective evaluation). A representative subset was selected from Set 1 and then adjusted for a SIM2 HDR monitor. Images were first cropped and scaled by a factor of two with a bilinear filter to fit their size to 944×1080 for side-by-side subjective experiments (details in Section 6), and then tone-mapped using display-adaptive TMO (Mantiuk et al 2008) to map the relative radiance representation of the images to an absolute radiance and color space of SIM2 HDR monitor. The regions to crop were selected by expert viewers in such a way that cropped versions were representative of the quality and the dynamic range of original images. Downscaling together with cropping approach was selected as a compromise, so that a meaningful part of an image can be shown on the SIM2 HDR monitor.

(TMOs) Since a TMO can be freely selected for encoding and its selection is not a part of JPEG XT specifications, we tested 5 different operators (the labels used in the results are given in parenthesis):

(*gamma*) a gamma clipping operator scales image values so that the average luminance value is equal to 1.0, then clamps all intensities to $[0, 1]$, and finally applies a gamma correction with an exponent of 2.2. This operator is the default setting in profile B.

(*drago03*) a global logarithmic tone-mapping operator (Drago et al 2003), which was found to give good compression performance (Mantiuk and Seidel 2008).

(*reinhard02*) a global version of the photographic operator (Reinhard et al 2002), which is a popular choice in many applications.

(*mai11*) a tone-mapping optimized for the best encoding performance in a backward-compatible scheme (Mai et al 2011).

(*mantiuk06*) a local operator with strong contrast enhancement (Mantiuk et al 2006b), which could be the most challenging case for a backward-compatible encoding scheme.

6 Subjective Evaluations

Subjective evaluations were conducted to achieve two main goals: a) to assess the perceptual visual quality of the three JPEG XT profiles discussed on this paper; and b) to validate and benchmark objective quality metrics so that they can be used for objective evaluations in Section 7.

Subjective evaluations were conducted at EPFL's Multimedia Signal Processing Group (MMSPG) test laboratory, which fulfills the recommendations for subjective evaluation of visual data issued by ITU-R (ITU-R BT.500-13 2012). The laboratory setup ensures the reproducibility of subjective test results by avoiding unintended influence of external factors. In particular, the laboratory is equipped with a controlled lighting system with a 6500 K color temperature, a mid gray color is used for all background walls and curtains, and the ambient illumination did not directly reflect off of the monitor. During the experiment, the background luminance behind the monitor was set to 20 lx.

To display the test stimuli, a full HD 47" SIM2 HDR monitor with individually controlled LED backlight modulation was used. Prior to subjective tests, following a warm-up phase of an hour, a color calibration of the HDR display was performed using the EasySolarPro software provided by SIM2 (SIM2 2015). The red, green, and blue primaries were measured at 1400 cd/m^2 level since the measurement probe (X-Rite i1Display Pro) is limited to a maximum value of 2000 cd/m^2.

In every session, three subjects assessed the displayed test images simultaneously. They were seated in an arc configuration, at a constant distance of 3.2 times the picture height, as suggested in recommendation (ITU-R BT.2022 2012).

[3] http://rit-mcsl.org/fairchild/HDR.html

[4] http://mmspg.epfl.ch/hdr-eye

Test Methodology The double-stimulus impairment scale (DSIS) Variant I methodology (ITU-R BT.500-13 2012) was selected, since this methodology is recommended for evaluating impairments and is typically used to evaluate compression algorithms. A five-grade impairment scale (1: *very annoying*, 2: *annoying*, 3: *slightly annoying*, 4: *perceptible, but not annoying*, 5: *imperceptible*) was used, since scales with a finer granularity are harder to handle for subjects and do not necessarily provide better resolving power.

Two images were presented in side-by-side fashion to reduce visual memory efforts by subjects. Due to the availability of only one full HD HDR monitor, each image was cropped and scaled (see the description of Set 2 in Section 5) to 944 × 1080 pixels with 32 pixels of black border separating the two images. One of the two images was always the reference (unimpaired) image. The other was the test image, which is a reconstructed version of the reference. Test images were created using the following procedure:

- Based on expert viewing, a TMO algorithm was chosen for each of the 20 images to produce the best visual quality. For 7 images, *reinhard02* TMO was selected and for 13 images *mantiuk06* was selected.
- For the selected tone-mapped version of each image, the JPEG quality parameter (q) was set to 4 different values such that they produce 4 different visual qualities based on expert viewing: very annoying, annoying, slightly annoying, and imperceptible (see Figure 3).
- The quality of the extension layer (Q) was then chosen for each profile in such a way that it would produce the same bit rate as that of the base layer. Such strategy resulted in a total of 12 (4 bit rates × 3 profiles) compressed versions for each HDR image (see Figure 3).
- A visual verification was then performed on HDR SIM2 monitor to confirm that 12 compressed versions of each HDR image cover the full quality scale from *very annoying* to *imperceptible*.

To reduce the effect of visual angle dependence, the participants were divided into two groups: the left image was always the reference image for the first group, whereas the right image was always the reference image for the second group. After the presentation of each pair of images, a six-second voting time followed. Subjects were asked to rate the impairments of the test images in relation to the reference image.

Test Design Before the experiment, a consent form was handed to subjects for signature and oral instructions were provided to explain their tasks. Additionally, a training session was organized allowing subjects to familiarize with the test procedure. For this purpose two images outside of the dataset were used. Five samples were manually selected by expert viewers for each image so that the quality of samples was representative of the rating scale.

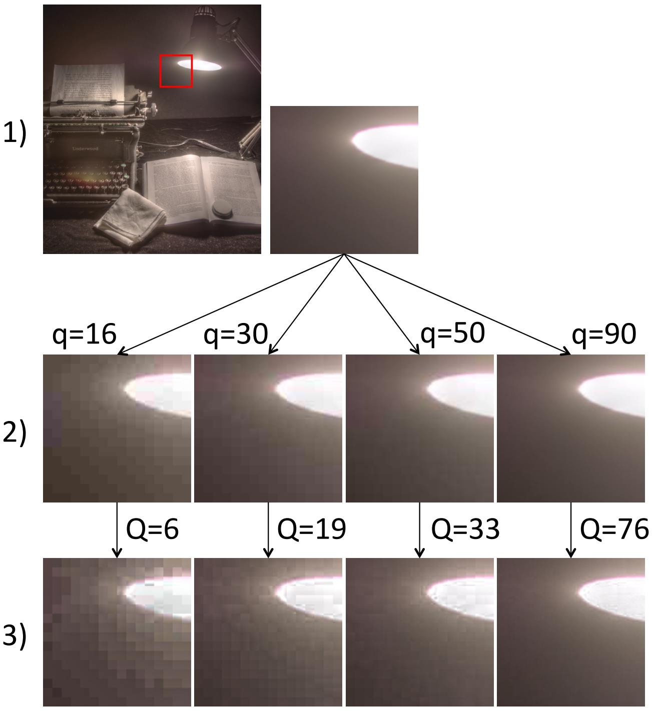

Fig. 3 Illustration of the test images creation process for LabTypewriter (Copyright 2006-2007 Mark D. Fairchild). 1) The TMO that produces the best visual quality is selected (*mantiuk06* in this case). 2) The tone-mapped image is encoded with JPEG at four different quality parameter (q) values such that they produce visual qualities corresponding to very annoying, annoying, slightly annoying, and imperceptible (q=16,30,50,90 in this case). 3) The HDR image is compressed with JPEG XT, using the base layer image and base layer quality parameter selected in 1) and 2), respectively. The quality parameter of the extension layer (Q) is set for each profile such that it produces the same bit rate as that of the base layer (Q=6,19,33,76 in this case for profile A). For printed representation, the compressed HDR images were tone-mapped with *mantiuk06*.

Since the total number of test samples was too large for a single test session, the overall experiment was split into 3 sessions of approximately 16 minutes each. Between the sessions, subjects took a 15-minute break. The test material was randomly distributed over the test sessions. To reduce contextual effects, the order of displayed stimuli was randomized applying different permutation for each group of subjects, whereas the same content was never shown consecutively.

A total of 24 naïve subjects (12 females and 12 males) took part in the experiments. Subjects were aged between 18 and 30 years old with an average of 22.1. All subjects were screened for correct visual acuity and color vision using Snellen and Ishihara charts, respectively.

Statistical Analysis The subjective scores were processed by first detecting and removing subjects whose scores deviated strongly from others. The outlier detection was applied to the set of results obtained from the 24 subjects and performed according to the guidelines described in Section 2.3.1 of Annex 2 of (ITU-R BT.500-13 2012). In this study, two outliers were detected. Then, the Mean Opinion Score (MOS) was computed for each test stimulus as the mean across scores by valid subjects, as well as associated 95% confidence interval (CI), assuming a Student's t-distribution of the scores.

6.1 Results

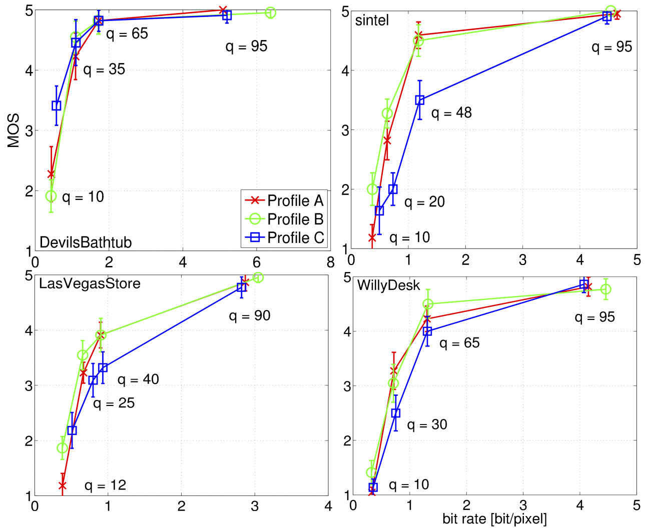

Fig. 4 Plots of the MOS at different bit rates for four of the images used in subjective tests for all three profiles. *Reinhard02* TMO was used for images reported in the first row and *mantiuk06* was used for others.

Figure 4 shows the plots of MOS at different bit rates for the three JPEG XT profiles, for a subset of the images used in subjective evaluations. The full set of plots is available in the supplemental material. In most cases, there is not sufficient statistical evidence to indicate differences in performance between profiles. However, at the lowest bit rates, profiles B and C outperform profile A on some contents. Likewise, for some contents, profile C shows lower performance at medium bit rates. Nevertheless, at the highest bit rates, all three profiles reach transparent quality.

After inspecting individual images we observed that profile A exhibits a lot of block coding artifacts in flat areas, similar to JPEG, but usually preserves colors, except at very low bit rates. Profile B suffers from color bleeding on areas of uniform colors, but exhibits less block coding artifacts than profile A. In addition, profile C performs better on flat uniform areas, but exhibits a checkerboard style color pattern on non-flat areas and introduces color noise near edges at low and medium bit rates, depending on content.

6.2 Benchmarking Of Quality Metrics

A second goal of subjective experiments was to evaluate how well an objective metric is capable of estimating perceived quality. To achieve this, the MOS obtained from subjective experiments are taken as ground truth and compared to predicted MOS values obtained from objective metrics. To compute the predicted MOS $\tilde{M}$, a regression analysis on each objective metric results O was performed on Set 2 using a logistic function as a regression model:

$$\tilde{M} = a + \frac{b}{1 + \exp\big(-c \cdot (O - d)\big)} \tag{6}$$

where a, b, c and d are the parameters that define the shape of the logistic fitting function and were determined via the least squares method.

Performance Indexes Performance indexes to assess the accuracy of objective metrics were computed following the same procedure as in (Hanhart et al 2013). In particular, the Pearson Linear Correlation Coefficient (PLCC) and the unbiased estimator of the Root-Mean-Square Error (RMSE) were used. The Spearman Rank Order Correlation (SROCC) coefficient and the outlier ratio (OR) were also used to estimate respectively the monotonicity and the consistency of the objective metric as compared with the ground truth subjective data. The OR is the ratio of points for which the error between the predicted and actual MOS values exceeds the 95% confidence interval of MOS values.

Statistical Analysis To determine whether the difference between two performance index values corresponding to two different metrics is statistically significant, two-sample statistical tests were performed on all four performance indexes. In particular, for the PLCC and SROCC, a Z-test was performed using Fisher z-transformation. For the RMSE, a F-test was performed, whereas a Z-test for the equality of two proportions was performed for the OR. No correction was applied to correct for the multiple comparisons. The statistical tests were performed according to the guidelines of recommendation (ITU-T P.1401 2012).

HDR Image Metrics A total of 12 objective quality metrics, including their variations, were tested. The main metrics were Peak Signal-to-Noise Ratio with maximum value equal to 1.0 ($PSNR$), Mean Root Square Error ($MRSE$), Signal-to-Noise Ratio (SNR), "Q" predictor in Visual Difference Predictor for HDR images HDR-VDP-2, version 2.2 ($HDRVDP\_Q$) (Mantiuk et al 2011), Structural-Similarity-Index ($SSIM$) and its multi-scale extension ($MSSSIM$) (Wang et al 2004). Since $PSNR$ and $MRSE$ were not intended to be used with HDR images, we included their variations, in which differences

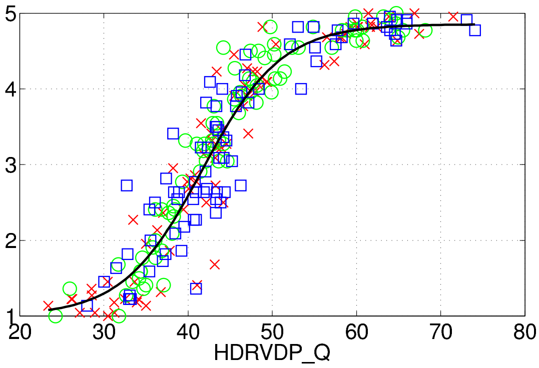

Fig. 5 Subjective versus objective evaluations results for the best performing metric.

are computed in the logarithmic domain (labels with *LOG_* prefix), or on Weber-like ratios: $(I_1-I_2)^2/(I_1^2+I_2^2)$ (*W_* prefix). Those values can be computed either on a luminance channel alone (*_Y* suffix) or on all three color channels (*_RGB* suffix). Finally, to adapt popular LDR metrics to HDR images, the Perceptually Uniform (PU) encoding (Aydın et al 2008) was used (*PU2* prefix).

Results The performance indexes computed on all contents at once are reported in Table 2. Results show that PSNR, LOG_PSNR_RGB, W_RMSE_RGB, SNR, RMSE and PU2PSNR_RGB perform poorly, as they have high RMSE and OR values. In contrast, HDR-VDP-2 applied on linear luminance values and SSIM and MSSSIM computed in the PU space are among the best metrics, with PLCC and SROCC values above 0.9. Figure 5 depicts the scatter plots of subjective versus objective results for these three metrics. For HDR-VDP-2, it can be observed that the data points do not deviate much from the logistic regression, which means that the prediction of the metric is consistent, as expressed by its relatively low OR. However, the deviation is higher for SSIM and MSSSIM, with OR values above 0.5.

The statistical analysis results are reported in Table 2. This analysis was performed on the performance indexes computed from 240 data points to discriminate small differences between two metrics. Results show that HDR-VDP-2 significantly outperforms other metrics, with the only exception being MSSSIM computed in the PU space. However, the OR of HDR-VDP-2 is statistically lower than that of PU2MSSSIM. Therefore, in the following part, we will consider only HDR-VDP-2.

The results show that commonly used metrics, such as PSNR, SNR, and MRSE, predict perceived quality of HDR content unreliably when computed on linear luminance values. However, performance is improved when the metrics are computed in a transformed perceptual space, e.g., the log or PU spaces. These results are supported by findings in (Hanhart et al 2014a) and (Valenzise et al 2014). However, the performance of these metrics improves if computed using only luminance (*_Y* suffix) than all RGB channels (*_RGB* suffix), which can be due to higher saliency of luminance artifacts compared to chromatic artifacts. The results also show that the best metric in terms of predicting perceived quality is HDR-VDP-2. It is different to stated in (Mantel et al 2014) that MRSE is the best quality predictor, but it could be because only six images were used in that study, and MRSE metric is content dependent.

7 Objective Evaluations

The results of subjective experiments are crucial in the selection of the right image quality metric and as a ground truth reference, but a subjective experiment alone cannot cover the entire space of parameters. Moreover, due the tedious nature of those experiments, only limited number of images can be tested, which makes the outcomes difficult to generalize. For that reason, we analyze the compression performance based on the results of the HDR-VDP-2, which was the best performing objective quality metric (see Table 2). Because of the scale of the required computation, the quality scores for 106 high-resolution images and in total 159 000 conditions were computed on an HPC cluster. The image quality computed for a range of base and extension layer quality settings may result in arbitrary bit rate, making the results difficult to aggregate. Therefore, the predicted quality values were linearly interpolated to find the HDR-VDP-2 Q-scores for each desired bit rate. This step was necessary to determine average performance and confidence intervals for all tested profiles. In the rest of this section, we will refer to predicted MOS, that means a MOS predicted from the HDR-VDP-2 Q-score based on the logistic function fitted to the subjective evaluation data (Figure 5).

First, we analyze how the setting of the base layer quality affects the compression performance. As shown in Figure 6 all profiles are relatively robust to the choice of the base layer quality. Interestingly, the performance of profiles A and C slightly improves with lower base-quality settings and for the bit-rates between 1 and 5 bit/pixel.

The performance of all three profiles is benchmarked in Figure 7. For fair comparison, we fixed the base layer quality setting at 80. In most applications the quality of the base layer cannot be tuned to achieve a slightly better performance for HDR compression. The backward-compatible layer must store an image of reasonable quality and setting around 80 provides such. The three colored plots demonstrate that profiles A and C achieve a similar compression performance at low bit rates, below 1.25 bit/pixel. At higher bit rate, profile C, which can encode up to 12-bit extension layer, shows a clear advantage. However, the quality gains for predicted MOS values above 4.5 are unlikely to be noticeable (about 50% of observers could not notice quality degradation at that level). Note that the plots show an average per-

Table 2 Accuracy (PLCC and RMSE), consistency (OR), and monotonicity (SROCC) indexes for each objective metric computed on all contents at once. Metrics whose performance indexes are underlined are considered statistically not significantly different. For example, according to PLCC, there is no statistical evidence to show performance differences between HDRVDP_Q and PU2MSSSIM, but they are statistically different from all other metrics.

(a) Pearson Linear Correlation Coefficient (PLCC).

LOG_PSNR_RGB	PSNR	SNR	PU2PSNR_RGB	W_RMSE_RGB	MRSE	W_RMSE_Y	PU2PSNR_Y	LOG_PSNR_Y	PU2SSIM	PU2MSSSIM	HDRVDP_Q
0.6548	0.6800	0.7128	0.7340	0.7386	0.7527	0.8812	0.8839	0.8881	0.9231	0.9447	0.9510

(b) Spearman Rank Order Correlation (SROCC).

LOG_PSNR_RGB	PSNR	SNR	PU2PSNR_RGB	MRSE	W_RMSE_RGB	PU2PSNR_Y	LOG_PSNR_Y	W_RMSE_Y	PU2SSIM	PU2MSSSIM	HDRVDP_Q
0.6004	0.6909	0.7150	0.7398	0.7524	0.7525	0.8844	0.8904	0.8915	0.9240	0.9499	0.9497

(c) Root-Mean-Square Error (RMSE).

LOG_PSNR_RGB	PSNR	SNR	PU2PSNR_RGB	W_RMSE_RGB	MRSE	W_RMSE_Y	PU2PSNR_Y	LOG_PSNR_Y	PU2SSIM	PU2MSSSIM	HDRVDP_Q
0.9487	0.9204	0.8805	0.8526	0.8466	0.8266	0.5941	0.5873	0.5770	0.4831	0.4133	0.3882

(d) Outlier ratio (OR).

PSNR	W_RMSE_RGB	SNR	MRSE	LOG_PSNR_RGB	PU2PSNR_RGB	W_RMSE_Y	PU2PSNR_Y	LOG_PSNR_Y	PU2SSIM	PU2MSSSIM	HDRVDP_Q
0.7625	0.7375	0.7208	0.7167	0.7000	0.6917	0.6208	0.5958	0.5833	0.5583	0.5250	0.3500

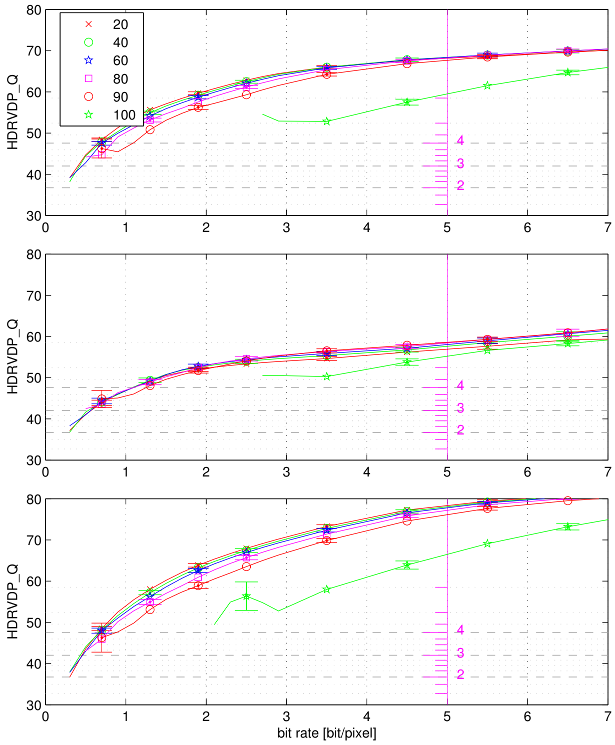

Fig. 6 The mean compression performance of each profile, averaged over 106 images and 5 TMOs, for different selection of the base layer quality: from 20 to 100. A higher HDRVDP_Q value denotes higher quality. The error bars denote 95% confidence intervals. The magenta scale shows MOS values corresponding to HDRVDP_Q predictions. From the top to the bottom the graphs correspond to Profile A, B and C respectively.

formance with high confidence (small 95% confidence interval), which is the result of averaging the results across the large number of tested images. The performance for individual images may vary substantially, as indicated by the box-plots in Figure 7. Note that these results are not directly comparable to those from the subjective experiment (Figure 4) because different base layer quality setting was used and image resolution was much higher.

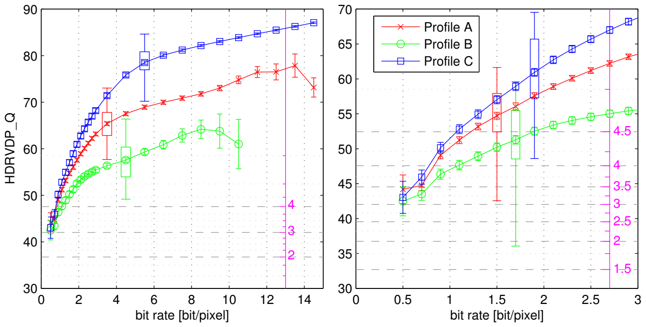

Fig. 7 The mean compression performance of each profile, averaged over 106 images and 5 TMOs. The quality of the base layer was fixed at 80. The error bars denote 95% confidence intervals. The box plots are drawn for three sample data points to visualize the distribution of the data: the boxes span from 25th to 75th percentile and the whiskers show 5th and 95 percentiles. The plot on the right focuses on the lower bit rates.

In Figure 8, we compare the performance of the three profiles with popular HDR image formats, including lossless OpenEXR and Radiance RGBE, and lossy JPEG 2000 and JPEG-XR (floating encoding). OpenEXR and Radiance offer lossless compression, however the loss happens when converting images to their internal pixel formats: 8-bit RGB channels and shared 8-bit mantissa (E) for Radiance RGBE; and 16-bit half-float (sign, 5-bit exponent, 10-bit mantissa) for OpenEXR. Note that our reference images were stored in 32-bit per-color channel, uncompressed PFM files. JPEG 2000 employs a lossy wavelet-based compression while JPEG-XR uses a two-stage frequency transform, combining the features of both DCT and wavelet transforms. HDR-VDP-2 did not detect any degradation in quality for all OpenEXR compression formats (HDRVDP_Q 100 is the highest qual-

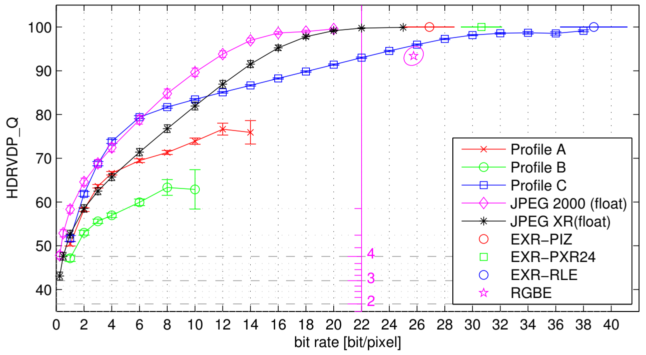

Fig. 8 The mean compression performance of each profile compared with popular near-lossless HDR image formats: OpenEXR using its three compression algorithms, and Radiance RGBE (.hdr). The base layer quality was fixed at 80. The ellipses denote 95% confidence interval.

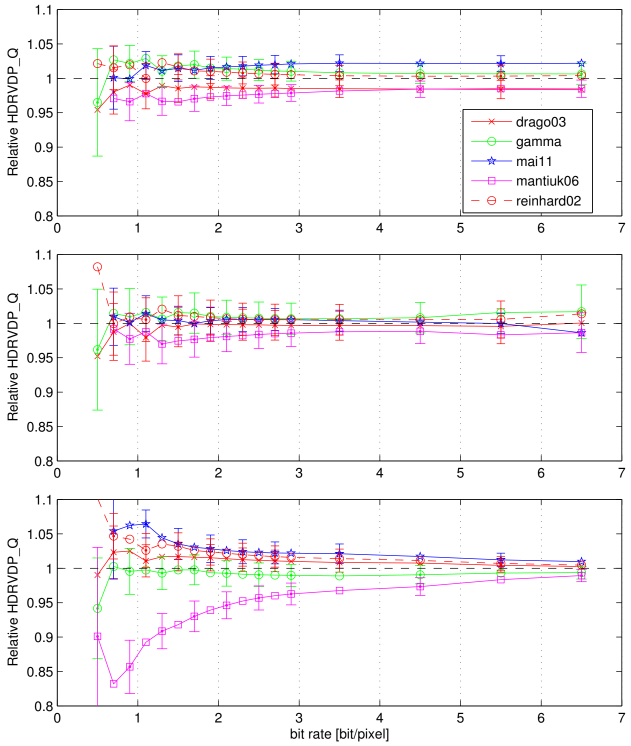

Fig. 9 The effect of a tone-mapping operator on the performance of each profile. The quality of the base layer was fixed at 80. The plot shows a relative change of quality as compared to an average performance for all operators. From the top to the bottom the graphs correspond to Profile A, B and C respectively.

ity), while small losses in quality were detected for Radiance RGBE. All those lossless formats preserve very high quality but require at least 26 bits per pixel (refer to EXR-* and RGBE data points in Figure 8). JPEG XT performs unexpectedly well when compared with other lossy compression methods. Below 10 bit/pixel, JPEG XT performs better than JPEG XR. Below 6 bit/pixel, the performance of JPEG XT is comparable to JPEG 2000, even though the former encodes an additional tone-mapped image and employs a standard DCT-based JPEG codec, rather than a more advanced compression algorithms found in both newer JPEG standards.

From Figure 7 and Figure 8, it can also be noted that all profiles can encode images at high quality (predicted MOS=4.5) with bit rates between 1.1 and 1.9 bit/pixel, while at least 27 bits per pixel are needed for the best performing OpenEXR PIZ compression (23× size reduction). Although "lossless" formats offer higher quality, the quality gain is unlikely to be noticeable at high predicted MOS values according to the correlations with subjective experiments (Section 6). The additional precision of these formats may be needed, however, if the content needs to be edited, tone-mapped or further processed. Only profile C offers encoding at precisions matching those offered by OpenEXR format. The bit rate of profile C for the same quality is slightly higher. However, profile C encodes additionally a backward-compatible base layer, which is missing in OpenEXR images.

We also analyze the influence of the tone-mapping selection on the compression performance. For better clarity, we show in Figure 9 the relative difference in quality as compared to the quality averaged over all TMOs (q/q_{mean}). A highly local operator *mantiuk06*, with strong detail enhancement, results in the lower compression performance for all profiles.

Profile C seems to be less robust to local operator (*mantiuk06* TMO) than the two other profiles. Not surprisingly, *mai11* TMO, which was explicitly optimized for backward-compatible HDR compression, results in improved performance. However, the gain between 2 and 5% is rather modest.

Next, we analyze how the bit rate is split between base and extension layers. Figure 10-top shows both bit rates for all three profiles and at four different quality settings of 60, 80, 90 and 100 for the base image. MOS values were mapped to corresponding HDRVDP_Q scores using the fitted logistic function. Up to moderately high quality settings (MOS=4.5, qbase=90), the extension layer occupies between 1/2 and 1/3 of the total bit rate. This ratio can drop to 10-20% when the highest JPEG quality setting is used for the base layer (qbase=100). This demonstrates that the overhead of HDR data is moderate.

So far we have analyzed an average compression performance. However, for many applications, the worst-case performance is even more important. For that reason, Figure 10-bottom shows the distribution of bit rates for the same quality criteria as discussed above. The plot shows that the total bit rate (base+extension) for a good quality compression (MOS=4.5, qbase=80) varies between 0.86 and 1.65 bit/pixel for 50% of the images, but in extreme cases can reach much higher rates. In

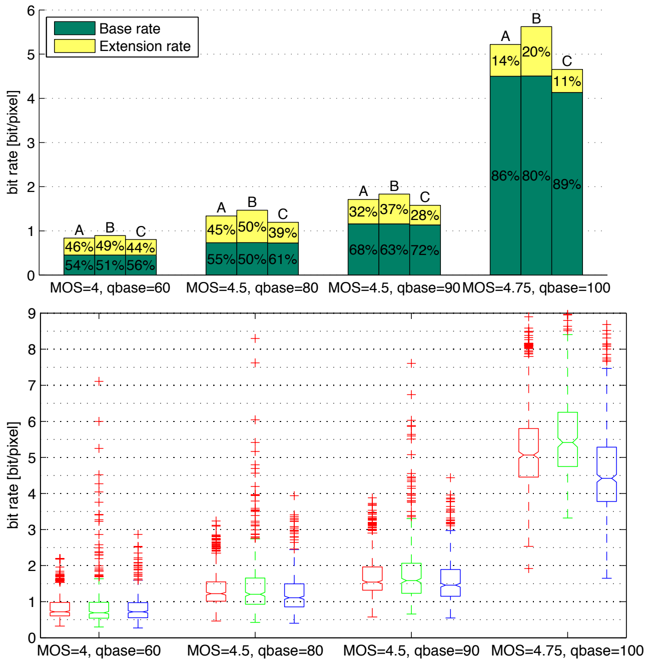

Fig. 10 Top: Bit rate of the base and extension layers for three predicted (by HDR-VDP-2) MOS quality levels and all three profiles (Profile A (red), B (green) and C (blue)). "qbase" is the quality setting used for the base layer. Bottom: The distribution of the total bit rate (base+extension) for the three profiles and the same quality criteria as in the top. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles and the whiskers extend to the most extreme data points not considered outliers.

terms of maximum bit rate for a fixed quality, profile B seems to be less robust. The highest bit/rate outliers produced by this profile could be responsible for its overall worse average performance.

8 Conclusion

The new upcoming standard called JPEG XT addresses an evident need for an efficient image format for HDR images. Even though there exist several standard that are capable of storing HDR images, such as JPEG 2000 and JPEG XR, they have not been widely adopted, most likely because they cannot be easily and cheaply integrated with the existing imaging infrastructure. In contrast to those, JPEG XT offers backward-compatibility with the most widely used 8-bit mode of ISO/IEC 10918/ITU Rec. T.81 (also known as JPEG) and minimizes investment in custom hardware. As consequence of this, the encoding or decoding hardware can be in fact designed based on a pair of existing JPEG coding chips, as shown in Figure 2, resulting in a minimal hardware change in the existing hardware infrastructure without influencing its real-time performances.

In this paper, we have presented the design philosophy, a high-level description and some technical details of the upcoming JPEG XT standard, followed by an extensive analysis of its performance under a wide range of coding conditions. The subjective evaluations demonstrated consistent HDR coding performance in a range of bit rates between 1 and 6 bits/pixel. The results of the subjective experiment were used for benchmarking of 12 image quality metrics and to select the most suitable metric for objective evaluations. The benchmarking showed that, in terms of predicting quality loss due to coding artifacts, simple metrics, such as PSNR, SNR, and MRSE computed in linear space are unsuitable for measuring perceptual quality of images compressed with JPEG XT, however, the prediction of these metrics improve when applied to the pixels converted in the perceptually uniform space. Also, HDR-VDP-2 provides the best performance as compared to other tested metrics. Objective evaluations using HDR-VDP-2 on 106 images demonstrated the robustness of the JPEG XT to the influence of its parameters: the quality for the base and extension layers and the tone mapping used for the base layer. Comparison to near-lossless and lossless existing formats show that JPEG XT is capable of encoding HDR images with bit rates varying from 1.1 to 1.9 bit/pixel that result in high estimated MOS (from objective measurements using HDR-VDP-2 metric) values of 4.5 out of 5 already. These results show that JPEG XT can achieve about 23 times reduction in file size, while maintaining comparable quality, when compared to lossless OpenEXR PIZ compression, as shown in Figure 1.

9 Acknowledgments

This work was partially supported by Ministry of Science and Innovation Subprogramme Ramon y Cajal RYC-2011-09372, TIN2013-47276-C6-1-R from Spanish government, 2014 SGR 1232 from Catalan government. EPFL authors acknowledge the Swiss National Foundation for Scientific Research (FN 200021-143696-1) which supported them in carrying out the work presented in this paper. The objective quality evaluation was possible thanks to High Performance Computing Wales, Wales national supercomputing service (hpcwales.co.uk).

References

Aydın TO, Mantiuk R, Myszkowski K, Seidel HP (2008) Extending quality metrics to full luminance range images. In: SPIE Human Vision and Electronic Imaging XIII, vol 6806

Chen M, Qiu G, Chen Z, Wang C (2006) JPEG compatible coding of high dynamic range imagery using tone mapping operators. In: Picture Coding Symposium (PCS), vol 1, pp 22–28

Drago F, Myszkowski K, Annen T, Chiba N (2003) Adaptive logarithmic mapping for displaying high contrast scenes. Computer Graphics Forum 22(3):419–426, DOI 10.1111/1467-8659.00689, URL http://www.blackwell-synergy.com/links/doi/10.1111%2F1467-8659.00689

Hanhart P, Korshunov P, Ebrahimi T (2013) Benchmarking of quality metrics on ultra-high definition video sequences. In: International Conference on Digital Signal Processing (DSP), pp 1–8, DOI 10.1109/ICDSP.2013.6622760

Hanhart P, Bernardo M, Korshunov P, Pereira M, Pinheiro A, Ebrahimi T (2014a) HDR image compression: a new challenge for objective quality metrics. In: QoMEX, pp 159–164

Hanhart P, Korshunov P, Ebrahimi T (2014b) Crowdsourcing evaluation of high dynamic range compression. In: SPIE Applications Of Digital Image Processing XXXVII, vol 9217

Husak W, Richter T (to appear) Information technology: Scalable compression and coding of continuous-tone still images, core coding system specification. International Organization for Standardization - ISO/IEC 18477-1

ITU-R BT2022 (2012) General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. International Telecommunication Union

ITU-R BT500-13 (2012) Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union

ITU-T P1401 (2012) Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. ITU

Korshunov P, Ebrahimi T (2013) Context-dependent JPEG backward-compatible high-dynamic range image compression. Optical Engineering 52(10)

Korshunov P, Hanhart P, Richter T, Artusi A, Mantiuk R, Ebrahimi T (2015) Subjective quality assessment database of HDR images compressed with JPEG XT. In: 7th International Workshop on Quality of Multimedia Experience (QoMEX)

Mai Z, Mansour H, Mantiuk R, Nasiopoulos P, Ward R, Heidrich W (2011) Optimizing a tone curve for backward-compatible high dynamic range image and video compression. IEEE Trans Image Processing 20(6):1558–1571, DOI 10.1109/TIP.2010.2095866

Mantel C, Ferchiu S, Forchhammer S (2014) Comparing subjective and objective quality assessment of HDR images compressed with JPEG XT. In: IEEE MMSP, pp 1–6, DOI 10.1109/MMSP.2014.6958833

Mantiuk R, Seidel HP (2008) Modeling a generic tone-mapping operator. Computer Graphics Forum 27(2):699–708

Mantiuk R, Krawczyk G, Myszkowski K, Seidel HP (2004) Perception-motivated high dynamic range video encoding. ACM Trans Graph 23(3):733, URL http://portal.acm.org/citation.cfm?doid=1015706.1015794

Mantiuk R, Efremov A, Myszkowski K, Seidel HP (2006a) Backward compatible high dynamic range MPEG video compression. ACM Trans Graph 25(3):713–723

Mantiuk R, Myszkowski K, Seidel H (2006b) A perceptual framework for contrast processing of high dynamic range images. ACM Trans Applied Perception 3(3):286–308, DOI 10.1145/1166087.1166095, URL http://portal.acm.org/citation.cfm?id=1166087.1166095

Mantiuk R, Daly S, Kerofsky L (2008) Display adaptive tone mapping. ACM Trans Graph 27(3):68, URL http://portal.acm.org/citation.cfm?id=1399504.1360667

Mantiuk R, Kim KJ, Rempel AG, Heidrich W (2011) HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans Graph 30(4):1, DOI 10.1145/2010324.1964935

Miller S, Nezamabadi M, Daly S (2013) Perceptual signal coding for more efficient usage of bit codes. SMPTE Motion Imaging Journal 122(4):52–59, DOI 10.5594/j18290, URL http://journal.smpte.org/cgi/doi/10.5594/j18290

Pattanaik S, Hughes C (2005) High-dynamic-range still-image encoding in JPEG 2000. IEEE Computer Graphics and Applications 25(6):57–64, DOI 10.1109/MCG.2005.133

Pennebaker WB, Mitchell JL (1992) JPEG Still Image Data Compression Standard. Van Nostrand Reinhold, New York

Pinheiro A, Fliegel K, Korshunov P, Krasula L, Bernardo M, Pereira M, Ebrahimi T (2014) Performance evaluation of the emerging JPEG XT image compression standard. In: IEEE MMSP, pp 1–6

Reinhard E, Stark M, Shirley P, Ferwerda J (2002) Photographic tone reproduction for digital images. ACM Trans Graph 21(3):267, DOI 10.1145/566654.566575

Richter T (2013a) Backwards compatible coding of high dynamic range images with JPEG. In: Data Compression Conference (DCC), pp 153–160, DOI 10.1109/DCC.2013.24

Richter T (2013b) On the standardization of the JPEG XT image compression. In: Picture Coding Symposium (PCS), pp 37–40, DOI 10.1109/PCS.2013.6737677

Richter T (2014) On the integer coding profile of JPEG XT. In: SPIE Applications Of Digital Image Processing XXXVII, vol 9217, DOI 10.1117/12.2060316

Richter T, Artusi A, Agostinelli M (to appear) Information technology: Scalable compression and coding of continuous-tone still images, HDR floating point coding. International Organization for Standardization - ISO/IEC 18477-7

Richter T, Husak W, Ninan A, Ten A, Jia W, Korshunov P, Ebrahimi T, Artusi A, Agostinelli M (to appear) Information technology: Scalable compression and coding of continuous-tone still images, extensions for high-dynamic range images. International Organization for Standardization - ISO/IEC 18477-2

Richter T, Ogawa S (to appeara) Information technology: Scalable compression and coding of continuous-tone still images, IDR integer coding. International Organization for Standardization - ISO/IEC 18477-6

Richter T, Ogawa S (to appearb) Information technology: Scalable compression and coding of continuous-tone still images, lossless and near-lossless coding. International Organization for Standardization - ISO/IEC 18477-8

Richter T, Schelkens P, Ishikawa T (to appear) Information technology: Scalable compression and coding of continuous-tone still images, box file format. International Organization for Standardization - ISO/IEC 18477-3

Richter T, Ten A, Artusi A (to appeara) Information technology: Scalable compression and coding of continuous-tone still images, conformance testing and evaluation. International Organization for Standardization - ISO/IEC 18477-4

Richter T, Ten A, Artusi A (to appearb) Information technology: Scalable compression and coding of continuous-tone still images, reference software implementation. International Organization for Standardization - ISO/IEC 18477-5

SIM2 (2015) SIM2 HDR display. URL http://www.sim2.com/

Spaulding K, Woolfe GJ, Joshi RL (2003) Using a residual image to extend the color gamut and dynamic range of an sRGB image. In: Proc. of IS&T PICS Conference, pp 307–314

Valenzise G, De Simone F, Lauga P, Dufaux F (2014) Performance evaluation of objective quality metrics for HDR image compression. In: Proc. SPIE 9217, Applications of Digital Image Processing XXXVII, pp 92,170C–92,170C

Wallace G (1992) The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics

38(1):xviii – xxxiv

Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: From error visibility to structural similarity. IEEE Trans Image Processing 13(4):600–612, DOI 10.1109/TIP.2003.819861

Ward G, Simmons M (2006) JPEG-HDR: a backwards-compatible, high dynamic range extension to JPEG. In: ACM SIGGRAPH 2006 Courses, DOI 10.1145/1185657.1185685, URL http://doi.acm.org/10.1145/1185657.1185685

Ward-Larson G (1998) LogLuv encoding for full-gamut, high-dynamic range images. Journal of Graph Tools 3(1):15–31, DOI 10.1080/10867651.1998.10487485