FINE-TUNING JPEG-XT COMPRESSION PERFORMANCE USING LARGE-SCALE OBJECTIVE QUALITY TESTING

Rafał K. Mantiuk [1], *Thomas Richter*[2] *and Alessandro Artusi*[3]

[1] Computer Laboratory, University of Cambridge, UK
[2] Computing Center, University of Stuttgart, Germany
[3] GiLab, Universitat de Girona, Spain

ABSTRACT

The upcoming JPEG XT standard for High Dynamic Range (HDR) images defines a common framework for the lossy and lossless representation of high-dynamic range images. It describes the decoding process as the combination of various processing tools that can be combined freely.

In this paper we analyze the coding efficiency of different decoding tools through a large scale objective quality testing using the HDR-VDP 2.2 objective metric. This evaluation is performed on a large database of 337 images, testing the effect of global and local tone mapping operators for various configurations, and for multiple combinations of quality parameters. The main findings are that using an inverse tone mapping operator for creating an HDR precursor image works well for global, but not for local operators, and that including refinement scans to increase the bit-depth of the extension layer provides substantial improvements for one of the encoding profiles and higher bit-rates.

Index Terms— JPEG XT, HDR Imaging, Tone Mapping

1. INTRODUCTION

In 2012, the JPEG Committee, formally known as ISO/IEC JTC1/SC29/WG1, issued a "call for proposals" to respond to the rapid increase interests in High Dynamic Range (HDR) imaging by industry. As result of this call, the JPEG XT standardization initiative has been started [1].

The standard defines a normative decoding procedure to reconstruct an HDR image from two JPEG regular codestreams, named the base layer which is visible to legacy decoders, and an extension layer, required for reconstruction of the HDR image. The standard does not, however, define the encoding procedure and leaves large freedom to the encoder for defining the necessary decoder configuration.

The main challenge of testing JPEG XT, as shown in [2], is that many factors influence its performance. The input of a JPEG XT *encoder* consists of two images: a base image visible to legacy decoders and an HDR image. Because two images are encoded, the compression performance does not depend on a single quality parameter only, but on two quantization settings for base and extension layers, and also on the choice of the tone-mapping operator (TMO) for generating the base image.

In this paper, our goal is to explore the whole space of possible configurations to achieve the best possible rate-distortion performance. To this end, our tests were conducted on a large dataset of 337 images, two TMOs, and multiple base and extension layer quality settings and 27 decoder configurations, resulting in over a million test cases in total. Clearly, this test corpus does not admit subjective testing, and for that reason we selected an objective metric, HDR-VDP 2.2 [3], for measurements. HDR-VDP 2.2 was selected as it was shown to provide the highest correlation to subjective scores according to several metric performance indices [2, 4].

2. RELATED WORK

Compression of high bit-depth still images, such as HDR images, was made possible by earlier standards such as JPEG 2000 [5] or JPEG XR [6], which can both encode high bit-depth integer and floating point data. However, both solutions are not compatible to the most wide-spread JPEG coding and their use is limited to some niche applications. Recognizing the lack of a JPEG-compatible standard for compression of HDR images, the JPEG Committee started JPEG XT as a new work item (formally, ISO/IEC 18477).

Studies evaluating JPEG XT on a limited set of images [7], a selection of profiles [8, 9], or on few objective metrics [10] appeared as early as 2014. Recently, an extensive study on the the performances of all profiles of JPEG XT have been provided [2, 4]. The latter works also demonstrated the high correlation between subjective scores and the HDR-VDP 2.2 objective quality index, indicating its usefulness for the evaluation of HDR content.

This work was partially supported .....

3. THE JPEG XT STANDARD

A JPEG XT image [1, 11, 12] consists of two JPEG images: the base layer visible to legacy decoders, and the extension (or residual) layer extending the precision from 8 bits to the necessary target precision for HDR representation. The bitrate for both of them can be controlled by quantization matrices which are, traditionally, configured by one quality parameter per layer. This results in a two-dimensional quality configuration space.

Metadata embedded in the JPEG XT file configures the decoder and defines how to merge both images into an output image. Profiles further constrain the allowable configurations to simplify implementation. While profile A configuration describes the HDR image as the pixel-wise product of the base layer times a scalar factor coming from the extension layer, a profile B decoder uses a quotient of base and extension layer components. Profile C uses a vector addition in the logarithmic space. For details, we refer to the literature [1].

4. TEST CONDITIONS

The image data set is an extended version of the one used in [2, 4]; it consist of a larger number of images (337 instead of 106), but at a lower resolution. HDR images were taken from two publicly available datasets: Fairchild's HDR Photographic Survey [13] and HDR-eye [14]. In both datasets, images have been obtained by merging multiple shots while varying the exposure. Whenever the resolution of an image exceeded the full HD resolution, it was both resized and cropped to at most 1920×1080 pixels, thus generating two separate test images.

To test the influence of the TMO on the compression performances, two different TMOs were chosen: A global tone-mapping designed for optimal performance of backward-compatible encoding (*mai11*) [15], and a local operator with strong contrast enhancement [16], which could be the most challenging case for a backward-compatible encoding scheme (*mantiuk06*). For measuring image quality, we selected HDR-VDP 2.2 [3] as previous studies [2, 4] demonstrated its high correlation to subjective performance.

The demo software available at `www.jpeg.org` was used as JPEG XT encoder as it allows configurations beyond the constraints of the JPEG XT profiles A, B and C.

5. RESULTS

Due to the large dataset (over one million test cases), quality scores were computed on an HPC cluster. Since the image quality is the function of both base and extension layer bit-rates, it would be difficult to analyze and visualize the results on 2D plots. Therefore, we report the result for the base and extension layer quality settings resulting in the highest HDR image quality for a given bit-rate. To compute the mean performance and confidence intervals for hundreds of tested images, the rate-distortion curves are linearly interpolated and sampled at fixed quality levels. We found that sampling at fixed quality levels rather than fixed bit-rates results in better overlap of the rate-distortion curves across images. Finally, for fair comparison, we use the same set of images for all configurations and data points shown on each plot. If all quality levels cannot be achieved for a certain image and for all configurations, that image is excluded from the average. For that reason, the results are averaged across a subset of images. Each plot reports the number of averaged images, "n", taken from the full set of 337.

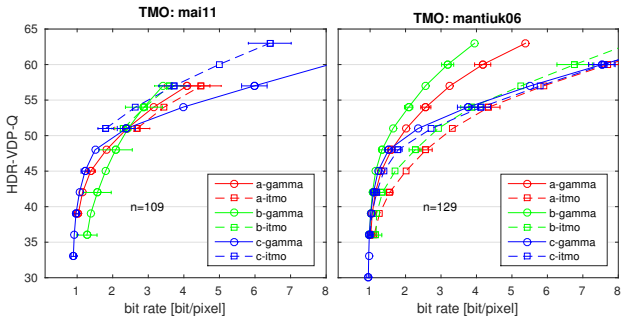


Fig. 1: Comparison between *gamma* and *itmo* prediction functions for global operator (left, *mai11*) and local (right, *mantiuk06*). n is the number of images over which the datapoints were averaged. The error bars denote 95% confidence intervals (plotted every second point to reduce clutter). A higher HDRVDP_Q value denotes higher quality. The profiles are labeled with colors: Profile A (red), B (green) and C (blue).

5.1. Inverse tone mapping function

We first test the influence of inverse tone-mapping function (*itmo*) on coding performance. The *itmo* is used to predict the pixel values in an HDR image based on the decoded base layer, thus it reduces the amplitude of values in the extension layer. Note that *itmo* is offered only in profile C and is an extension of profiles A and B.

The results in Figure 1 show a substantial difference in performance between the two tested TMOs and profiles. Contrary to expectations, *itmo* does not improve coding performance for profile A and B, and even reduces performance when the base layer is generated with the local TMO (*mantiuk06*). However, for profile C coding performance is significantly improved with *itmo* for global operator (*mai11*) and slightly reduced for the local TMO. Therefore, *itmo* seems to be beneficial only for Profile C and global TMOs.

Generating the base image through a local TMO clearly results in less-compressible images as the local TMO tends to boost high spatial frequencies. However, *itmo* computed

(a) Base non-linearity is $\gamma : 2.2$ (b) Base non-linearity is iTMO

Fig. 2: Residual images of a profile C configuration and a base image generated by a local operator. Left for a base-map with fixed gamma, right for an inverse TMO. Both are fully in profile C.

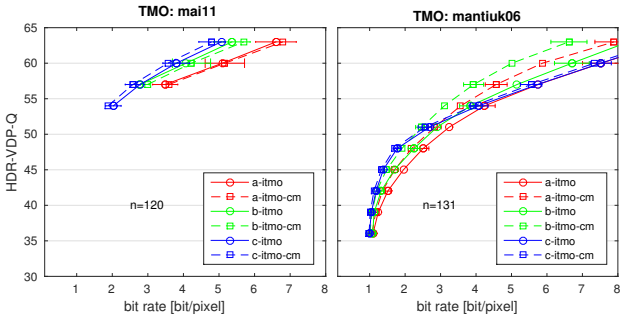


Fig. 3: The comparison between *itmo* reconstruction function computed using means (*itmo*, continuous lines) and medians (*itmo-cm*, dashed lines).

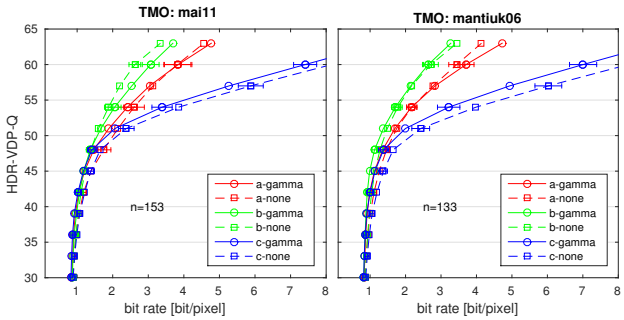


Fig. 4: Comparison of the prediction function set to $\gamma = 2.2$ (labeled *gamma*, continuous lines) and to the identity function (labeled *none*, dashed lines).

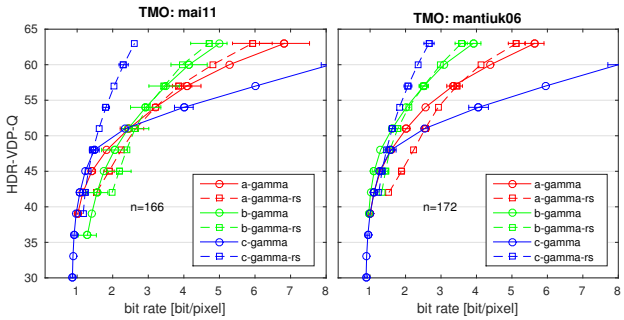


Fig. 5: Comparison of the configurations without a refinement scan (*gamma*, continuous lines) and with a refinement scan (*gamma-rs*, dashed lines).

for a given image should result in better rather than worse compression performance. This discrepancy is explained in Fig. 2, which shows the residual image for the profile C configuration with a base image generated by the local operator: On the left, the residual image for a pre-cursor image given by the $\gamma = 2.2$ mapped base image, on the right the residual image if the pre-cursor image is generated by *itmo*. The overall amplitudes of the left residual, produced without *itmo*, are larger but are more regular and hence easier to encode. The iTMO non-linearity generates smaller amplitudes but a much less regular residual image. Since *itmo* is computed independently per-pixel, it can easily introduce additional high frequency variations in the residual image.

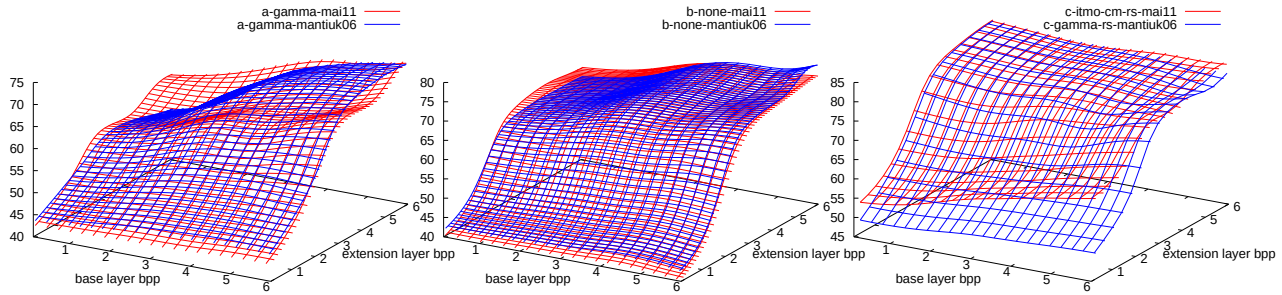
In the second experiment, we confined ourselves to *itmo*, but varied the algorithm used to obtain it from the LDR/HDR image pair. In a general case, for all pixels with a fixed luma value in an LDR image, the corresponding pixels in the HDR image have different luminance values. To predict the most likely HDR luminance value from a LDR luma value, one can compute the mean (in the logarithmic domain) of all corresponding HDR luminance values. This is the default configuration, denoted as *itmo*. But the prediction can be also computed as median (center of mass), and we denote it as *itmo_cm*.

Figure 3 shows the results for median (continuous) and center of mass (dashed) *itmo* configurations, both for a global TMO (left) and a local TMO (right). As seen there, the center of mass algorithm provides better performances in all cases, in particular in the case of the local TMO. Clearly, this algorithm minimizes the expectation value of the error residual to be encoded in the extension layer, and hence helps by reducing the amplitudes of the residual.

So far we have compared fitted *itmo* to a fixed *gamma* function used to predict HDR pixel values from the decoded base layer. But the prediction function could also be set to the identity function (gamma=1), which we denote as *none* in Figure 4. The lack of gamma non-linearity degrades the performance for most profiles, except for profile B and the global operator. No gamma is also beneficial for high bitrates for profile A.

5.2. Refinement scans

In the third experiment, we studied the impact of an additional extension mechanism of JPEG XT denoted as “refinement scans”. Refinement coding extends the coding precision in the DCT domain thanks to a coding mechanism that is closely related to the progressive coding mode of the legacy JPEG standard. It extends the coding precision to 12-bits in



(a) Profile A (b) Profile B (c) Profile C

Fig. 6: Overall performance of all three profiles averaged over all images, in their best configuration, as a function of base and extension bit-rates. In red, with global TMO, in blue with local TMO. Note that the scales are not identical.

.

the spatial domain. The most significant bits of the quantized DCT coefficients are encoded by a regular JPEG coding mode, forming the codestream that legacy applications can interpret. The least significant bits are encoded with the so-called successive approximation scan, which is a part of the progressive coding mode also defined in the legacy JPEG standard [1, 11].

Refinement coding can be applied both to the base layer and/or to the extension layer; we here constrained ourselves to testing the latter case only. The performances of all profiles when using a refinement scan option are benchmarked in Figure 5. At higher bit rates refinement scan (dashed lines) help to improve the quality performances most notably for profile C. For profiles A and B refinement scans are extensions outside of the profile while they are part of profile C.

5.3. Overall performance analysis

The above experiments indicate that the most suitable configuration is to use a *gamma* prediction function for profile A, and the identity function (gamma=1) for profile B, regardless of which tone mapping operator has been applied. For profile C, it is advisable to enable refinement scans to reach higher performance unless very low bit-rates are required. The choice of the prediction function for profile C, however, depends on tone mapping. For global TMOs, an *itmo* works best, for local TMOs, a *gamma* improves performance as it produced a more regular residual image (cf. Fig. 2).

The plots in Fig. 6 show now the overall quality of the three profiles as a function of the base and extension layer bit-rate, averaged over the entire dataset. At first, it seems surprising that the average image quality depends so little on the rate of the base layer. A closer inspection shows that the variance between images is quite high, which diminishes the dependency on the base layer rate when average data is presented. This is especially true for profile B. For profile C, we found that increasing the base rate at constant extension bit-rate (*not* at constant extension q parameter!) does not always improve the overall image quality; this is because a finer base quantization injects higher frequencies into the residual layer which reduce its coding performance, and hence increase the extension layer rate at constant extension q. It is, however, the only profile that allows lossy to lossless coding.

6. CONCLUSIONS

In this paper we presented fine-tuning of the JPEG XT standard using large-scale objective quality testing. The use of an *itmo* prediction function has been found useful for improving the quality performance for Profile C, but only when a global TMO is used. The *itmo* configuration lowers the performance compared to a simple gamma for all profiles when the base image is generated by a local TMO. We have also observed that if an *itmo* prediction function is used, it should be computed as the medians of the LDR-to-HDR mapping. This observation is consistent and independent of the decoder configuration.

Refinement scans over the extension image substantially help to improve the performance of profile C for higher bit depths, but they do not improve quality as much for an extension of profiles A and B because both include an equivalent scaling mechanism for the residuals. However, refinement scans also increase the overhead for lower bitrates and hence should not be used for high compression, i.e. low quality.

7. ACKNOWLEDGEMENTS

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. Thomas Richter thanks the Computing Center of the University of Stuttgart for supporting this work. The objective quality evaluation was possible thanks to High Performance Computing Wales, Wales national supercomputing service (hpcwales.co.uk).

8. REFERENCES

[1] T. Richter, T. Bruylants, P. Schelkens, and T. Ebrahimi, "The JPEG XT Suite of Standards: Status and Future Plans," in *Proc. SPIE 9599, Applications of Digital Image Processing XXXVIII*, 2015.

[2] A. Artusi, R. Mantiuk, R. Thomas, H. Philippe, K. Pavel, A. Massimiliano, T. Arkady, and E. Touradj, "Overview and evaluation of the JPEG XT HDR image compression standard," *Real Time Image Processing Journal*, 2015.

[3] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high dynamic range and standard images," *Journal of Electronic Imaging*, vol. in print, 2015.

[4] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, "Subjective quality assessment database of HDR images compressed with JPEG XT," in *7th International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015.

[5] ITU, "Recommendation ITU-T Rec. T.800, Information technology: JPEG 2000 image coding system: Core coding system," International Telecommunications Union, 2002.

[6] ——, "Recommendation ITU-T Rec. T.832 ISO/IEC 29199-2, Information technology: JPEG XR image coding system Image coding specification," International Telecommunications Union, 2012.

[7] A. Pinheiro, K. Fliegel, P. Korshunov, L. Krasula, M. Bernardo, M. Pereira, and T. Ebrahimi, "Performance evaluation of the emerging JPEG XT image compression standard," in *IEEE MMSP*, 2014, pp. 1–6.

[8] C. Mantel, S. Ferchiu, and S. Forchhammer, "Comparing subjective and objective quality assessment of HDR images compressed with JPEG XT," in *IEEE MMSP*, 2014, pp. 1–6.

[9] P. Hanhart, M. Bernardo, P. Korshunov, M. Pereira, A. Pinheiro, and T. Ebrahimi, "HDR image compression: a new challenge for objective quality metrics," in *QoMEX*, 2014, pp. 159–164.

[10] G. Valenzise, F. De Simone, P. Lauga, and F. Dufaux, "Performance evaluation of objective quality metrics for HDR image compression," in *Proc. SPIE 9217, Applications of Digital Image Processing XXXVII*, 2014, pp. 92 170C–92 170C.

[11] A. Artusi, R. Mantiuk, R. Thomas, K. Pavel, H. Philippe, E. Touradj, and A. Massimiliano, "Jpeg xt: A compression standard for hdr and wcg images," *IEEE Transaction on Signal Processing Magazine*, 2016.

[12] T. Richter, W. Husak, , N. Ajit, T. Arkady, P. Korshunov, T. Ebrahimi, A. Artusi, M. Agostinelli, S. Ogawa, P. Schelkens, T. Ishikawa, and T. Bruylants, "JPEG XT information technology: Scalable compression and coding of continuous-tone still images." [Online]. Available: http://jpeg.org/jpegxt/index.html

[13] M. Fairchild, "Fairchild's hdr photographic survey," 2008. [Online]. Available: http://rit-mcsl.org/fairchild/HDR.html

[14] P. Korshunov and T. Hanhart, Philippe ad Ebrahimi, "Epfl's dataset of hdr images," 2015. [Online]. Available: http://mmspg.epfl.ch/hdr-eye

[15] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, "Optimizing a tone curve for backward-compatible high dynamic range image and video compression," *IEEE Trans. Image Processing*, vol. 20, no. 6, pp. 1558–1571, 2011.

[16] R. Mantiuk, K. Myszkowski, and H. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Applied Perception*, vol. 3, no. 3, pp. 286–308, 2006.