Uniform Color Space based High Dynamic Range Video Compression - supplementary material

R. Mukherjee[1,4], K. Debattista[1], T. Bashford-Rogers[2], M. Bessa[3,4] and A. Chalmers[1]

[1]University of Warwick, UK
[2]University of the West of England, UK
[3]Universidade de Trás-os-Montes e Alto Douro, Portugal
[4]INESC TEC, 4200-465 Porto, Portugal

I. Introduction

This document contains some of the contents which could not be accommodated in the manuscript due to page constraints. Section II provides some of the results from the PTF evaluation which could not be included in the main manuscript due to page constraints. Section III provides a brief discussion on the metadata used during compression and decompression of HDR frames using the proposed algorithm. Section IV include the mean and interpolated RD characteristics from the overall objective evaluation of the five algorithms e.g. *hdrv*, *fraunhofer*, *pq*, *bbc-hlg* and the *proposed* algorithm. Section V provides a brief discussion on the subjective evaluation and presents the correlation results from non-parametric tests such as Kendall's Tau and Spearman's Rho rank correlation. Finally, we present a tone-mapped thumbnail for each of the 39 sequences used for the objective evaluation.

II. Perceptual Transfer Functions (PTFs)

In this section, we provide some of the results from the evaluation of the five PTFs, including the proposed PATF. As discussed in Section IV-C1 of the main manuscript, the proposed PATF was evaluated against the four existing state-of-the-art PTFs following *Pipeline-B* of the evaluation methodology shown in *Figure 6* of the main manuscript. The main manuscript contains two of the mean results obtained from the PTF evaluation. However, as mentioned in the main manuscript, the mean RD graphs do not provide a detailed understanding of the quality of each PTF as the mean encoding bitrates and output quality is largely dependent on the encoded sequence. Therefore, to provide fair comparison results across all sequences, we provide interpolated results where the same set of sequences are averaged for each PTF and data is provided for a fixed set of bitrates (quality levels). The RD characteristics also demonstrate the variation in image reconstruction quality across the set of 39 sequences with 95% confidence interval bars as shown in Figure 1.

A. PTF evaluation analysis

The interpolated RD characteristics of the five PTFs preseted in Figure 1 establish that the proposed PATF outperforms existing solutions. Moreover, as shown and mentioned in the main manuscript, the cone-response function (GCRM) and Ferwerda t.v.i are closely matched with the proposed PTF. Therefore, amongst the existing solutions, GCRM and Ferwerda's t.v.i can also be used in conjunction with the proposed algorithm instead of the proposed PATF, albeit at the cost minor quality degradation. Similar to the results presented in the main manuscript, the adaptive logarithmic function and GDF are not ideally suited for the best reconstruction of decoded HDR frames.

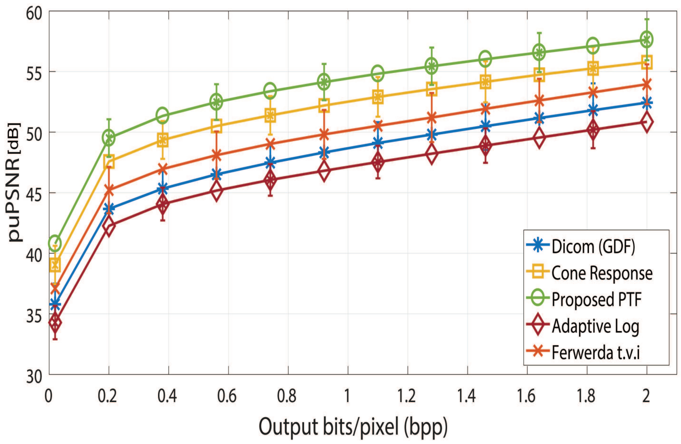
III. Metadata information (of the proposed algorithm)

As mentioned in Section IV-E of the manuscript, the proposed algorithm generates metadata which contains; a) the normalization factor ν which is typically the maximum pixel value of the input HDR frame, b) the multiplying factor ψ which is used to scale the normalized intensity channel say $I \in [0,1]$ to a range suitable for the chosen PTF (say $I' \in [0.05, 4000]$ for DICOM GDF), c) the minimum and maximum values of the chroma channels prior to scaling and d) the power value λ applied to each chroma channel during the application of the error minimization function (see Section IV-D) of the main manuscript.

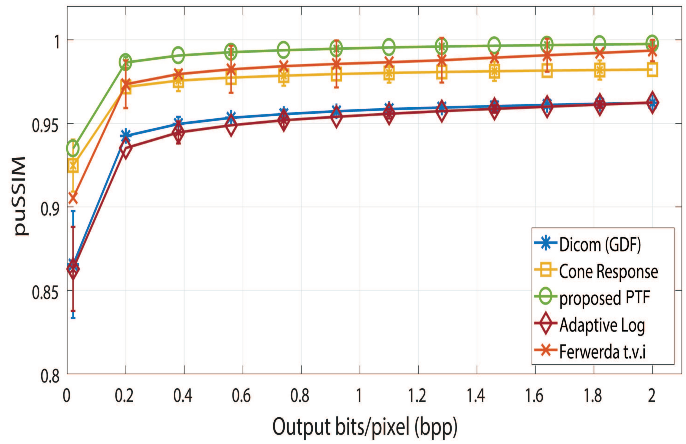
The metadata information is stored for every frame in a lookup table (LUT) and the information is used later for reconstruction of the HDR frames. Table I provides a visual description of an example LUT.

This metadata is vital for accurate reconstruction of HDR frames. However, the reliability on auxiliary information is not always desired for compression and transmission purposes as corruption of metadata would lead to faulty HDR reconstruction. Therefore, we also propose an alternate solution which eliminates the requirement for auxiliary metadata, albeit at the cost of reconstruction quality. To that end, we assume a few constants such as a) the intensity channel values are to be scaled such that $\mathrm{I}' \in [10^{-4}, 10^{4}]$, irrespective of the PTF applied to map the scaled intensity values to a 10-bit JND scale, b) the accurate scaling of chroma pixel values where $P, T \in [-1, 1]$ are to be replaced by a straightforward *addition and multiplication* routine to map them to a $[0, 1]$ range and c)the error minimization function is replaced with a fixed non-linear gamma encoding such that $\gamma = \frac{1}{2.2}$.

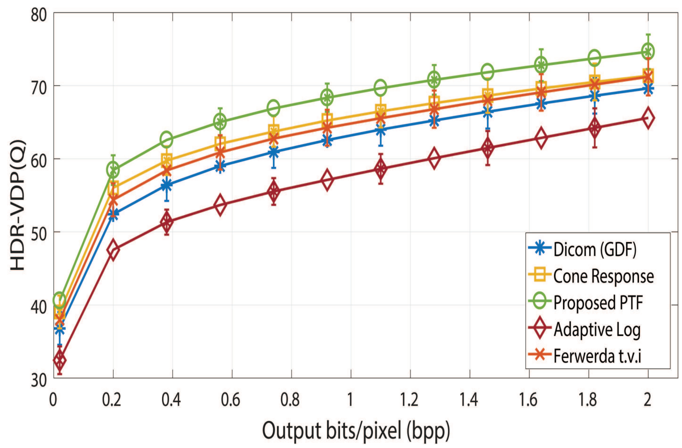
Upon objective evaluation of the proposed algorithm, with the assumed constants, it was found that the image recon-



(a) RD characteristics of the five PTFs: puPSNR - (higher is better)



(b) RD characteristics of the five PTFs: puSSIM - (higher is better)



(c) RD characteristics of the five PTFs: HDR-VDP(Q) - (higher is better)

Fig. 1: Interpolated RD characteristics of the five PTFs - averaged over 39 sequences. The evaluations follow Pipeline B mentioned in the main manuscript.

FrameNo	ν	I_{scale}	P_{min}	P_{max}	T_{min}	T_{max}	λ_P	λ_T
00000	4658	4000	-0.567	0.892	-0.124	0.589	0.899	0.967
00001								
.....								
00149								

TABLE I: Example metadata information look-up table.

struction quality was slightly lower compared to the metadata solution as presented in the main manuscript. However, these modifications facilitates a minor increase the overall performance of the proposed algorithm.

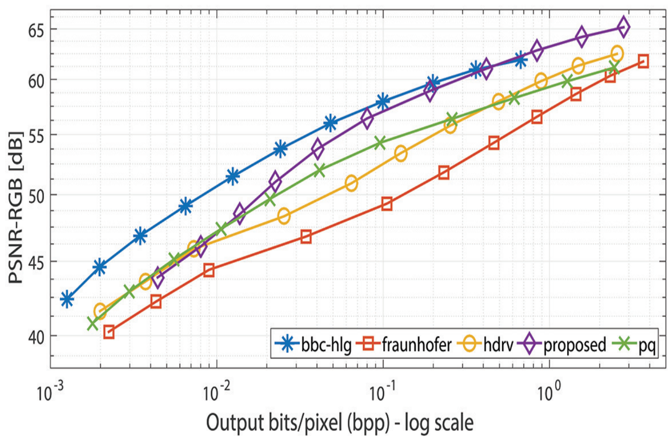
IV. Results

In this section, we present a detailed set of results obtained from the objective evaluation of the proposed algorithm compared to existing state-of-the-art solutions. These include the mean RD graphs (see Figure 2) from the seven QA metrics used in this evaluation (apart from the ones shown in the main manuscript) as shown in Figure 2.

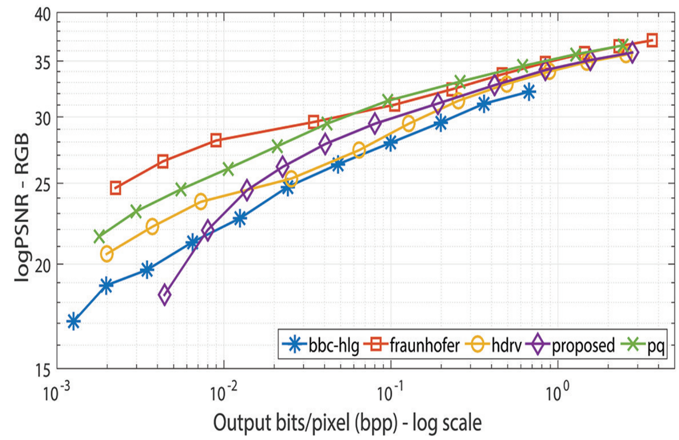
As mentioned earlier, we also present the set of interpolated RD characteristics which ensure that the RD curves do not overlap and demonstrate the mean output quality along with the variation in either image quality at fixed bitrates or variation in encoding bitrates at fixed quality levels as shown in Figures 3 and Figure 4, respectively.

V. Subjective Evaluation

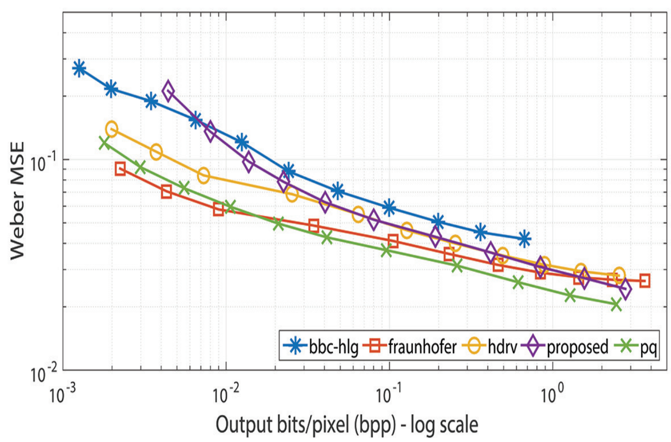
In this section, we provide some further details of the subjective evaluation. As mentioned in the main manuscript, six sequences were short listed from the set of the 39 sequences. The selection ensured that the dynamic ranges of the selected sequences have significant variation. A variation in capturing technique was also ensured whereby some sequences were captured using either the Spheron VR camera or the ARRI



(a) RD characteristics PSNR-RGB (higher is better)



(b) logPSNR-RGB (higher is better)



(c) Weber MSE (lower is better)

Fig. 2: Mean RD characteristics of the five algorithms - averaged over 39 sequences.

Alexa. One of the sequences was from the open movie “Tears of Steel” by Blender foundation which was partly artificially rendered and partly captured. Figure 5 provides a thumbnail of each of the six sequences used in the subjective evaluation.

The six short listed sequences were shortlisted and two quality levels $Q_1 = 0.09bpp (\approx 6Mbps)$ and $Q_2 = 0.24bpp (\approx 15Mbps)$ were chosen such that they represent lower and higher quality, respectively. The six sequences were encoded at different the specified average output bitrates and the objective results were stored. Subsequently, the output from the algorithms for each of the six sequences along with the uncompressed *reference* was used for a rating-based experiment as mentioned in the main manuscript.

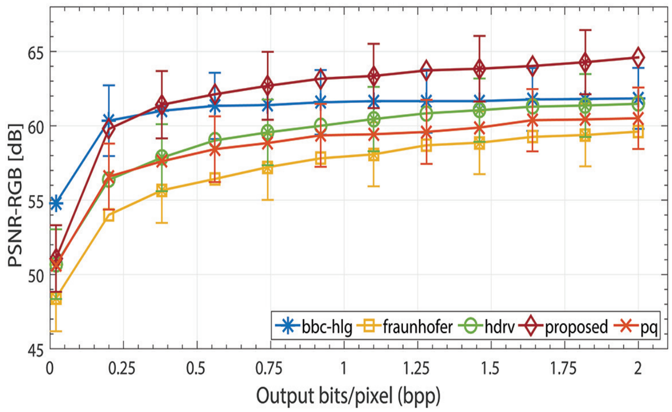
Following, the subjective evaluation, the results were folded across the quality levels averaged per scene. Finally, the results were averaged per algorithm. Similar procedure was conducted for the objective results at the specified output bitrates and subsequently, the correlation between the results were calculated using parametric and non-parametric tests such as Pearson’s, Kendall Tau and Spearman’s Rho rank correlation. Since, we have presented the Pearson’s correlation results in the main manuscript, we present the other two results in Tables II and III, respectively.

	puPSNR	puSSIM	HDR-VDP	HDR-VQM	Subjective
puPSNR	-	1.000**	1.000**	1.000**	.800
puSSIM	1.000**	-	1.000**	1.000**	.800
HDR-VDP	1.000**	1.000**	-	1.000**	.800
HDR-VQM	1.000**	1.000**	1.000**	-	.800
Subjective	.800	.800	.800	.800	-

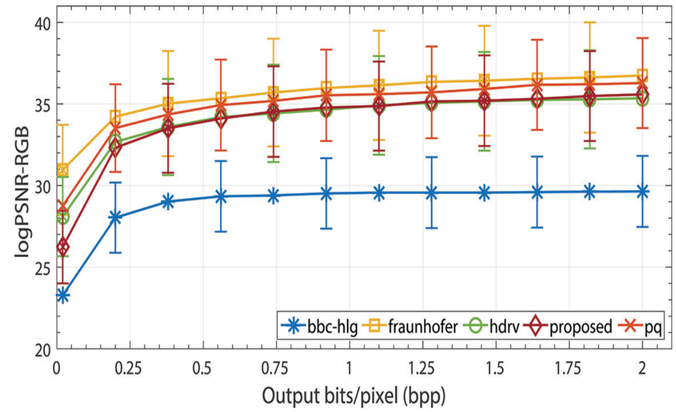
TABLE II: Kendall’s Tau correlation results. ‘*’ represents significance at $p < 0.05$ and ‘**’ represents significance at $p < 0.01$.

	puPSNR	puSSIM	HDR-VDP	HDR-VQM	Subjective
puPSNR	-	1.000**	1.000**	1.000**	.900*
puSSIM	1.000**	-	1.000**	1.000**	.900*
HDR-VDP	1.000**	1.000**	-	1.000**	.900*
HDR-VQM	1.000**	1.000**	1.000**	-	.900*
Subjective	.900*	.900*	.900*	.900*	-

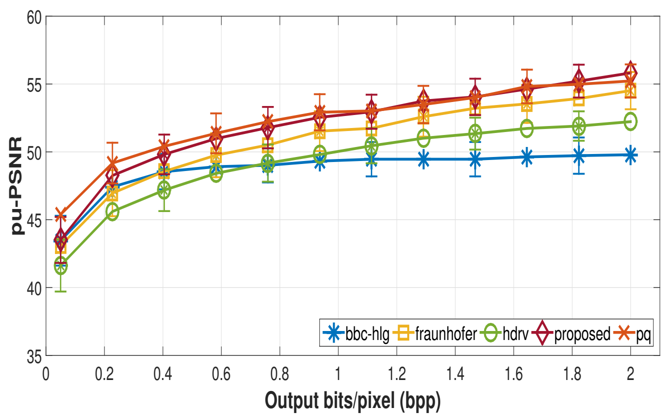
TABLE III: Spearman’s rho rank correlation results. ‘*’ represents significance at $p < 0.05$ and ‘**’ represents significance at $p < 0.01$.



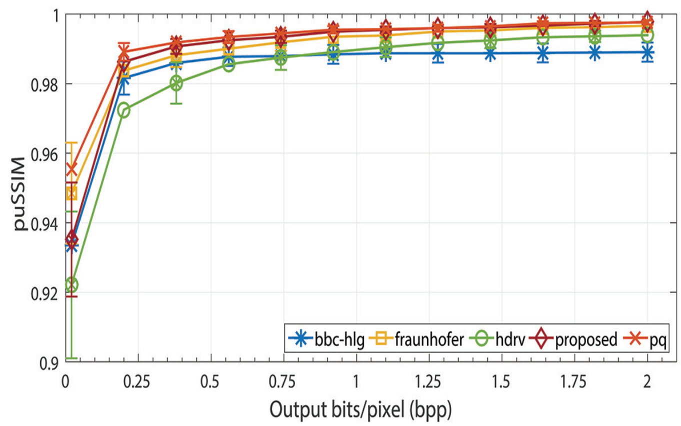
(a) RD characteristics PSNR-RGB (higher is better)



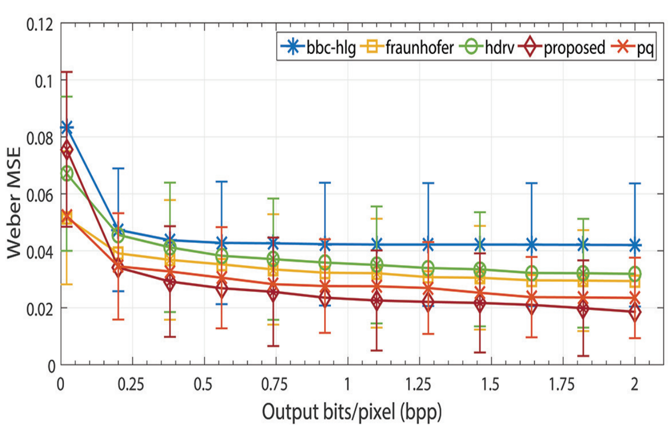
(b) logPSNR-RGB (higher is better)



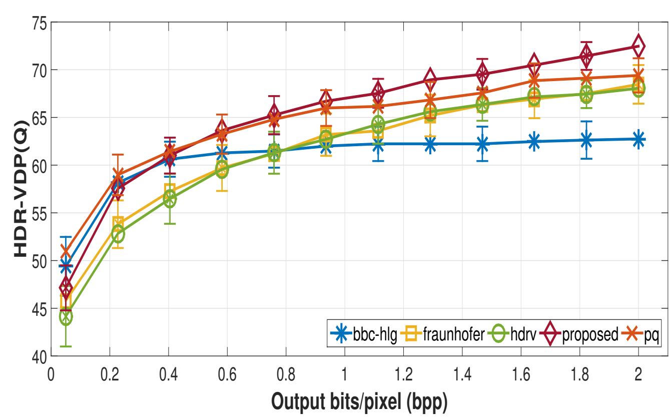
(c) puPSNR (higher is better)



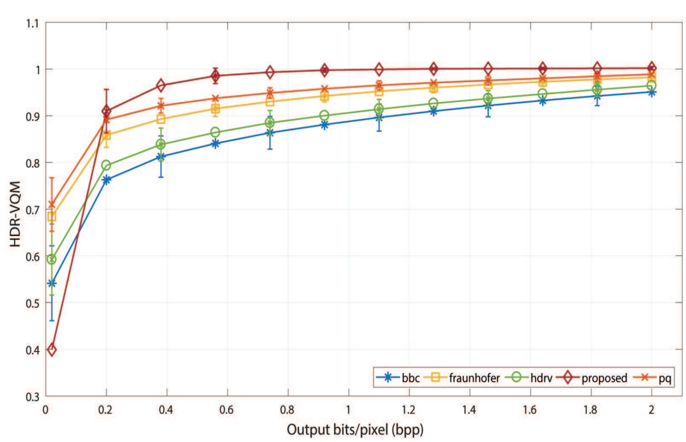
(d) puSSIM (higher is better)



(e) Weber MSE (lower is better)

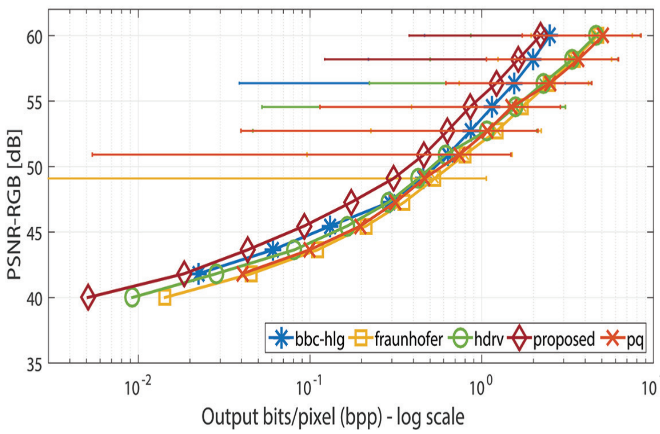


(f) HDR-VDP(Q) (higher is better)

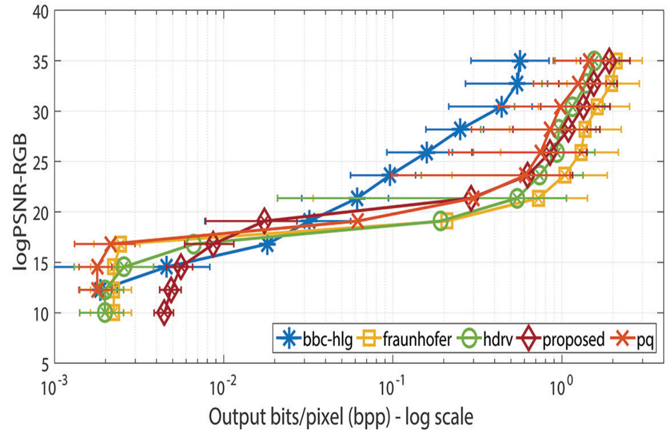


(g) HDR-VQM (higher is better)

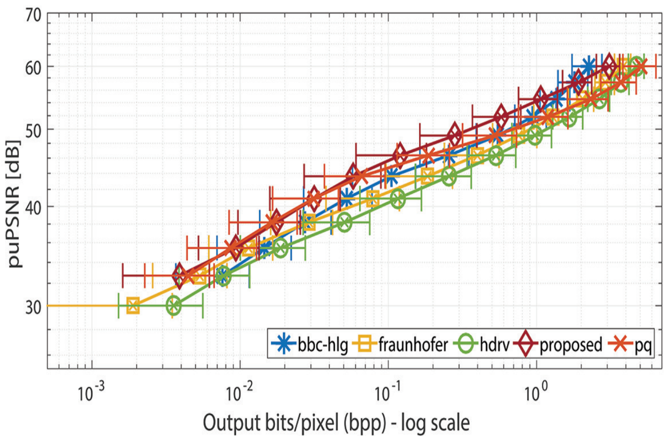
Fig. 3: Interpolated RD characteristics of the five algorithms at fixed bitrates (exhibiting variation in image quality) - averaged over 39 sequences.



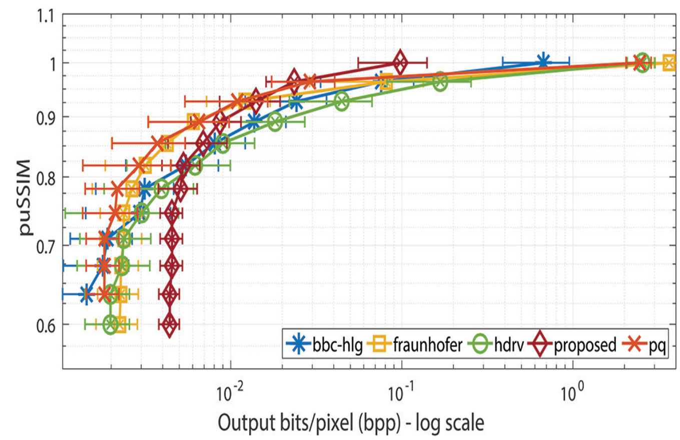
(a) RD characteristics PSNR-RGB (higher is better)



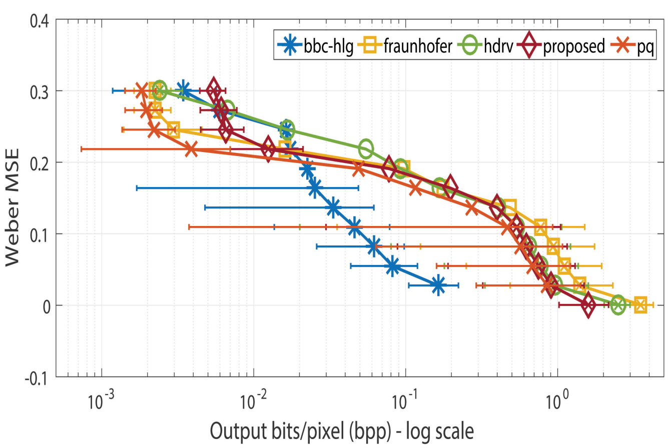
(b) logPSNR-RGB (higher is better)



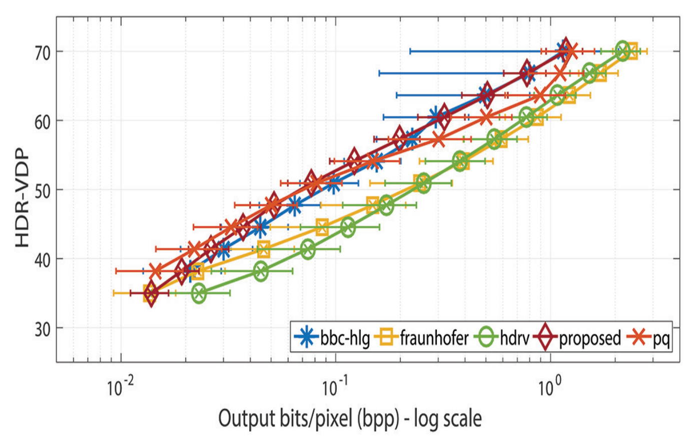
(c) puPSNR (higher is better)



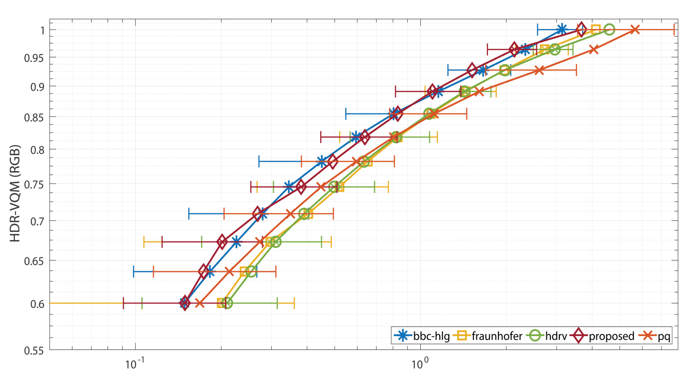
(d) puSSIM (higher is better)



(e) Weber MSE (lower is better)



(f) HDR-VDP (higher is better)



(g) HDR-VQM (higher is better)

Fig. 4: Interpolated RD characteristics of the five algorithms at fixed quality levels (exhibiting variation in bitrate) - averaged over 39 sequences.

(a) Welding - 21 stops (b) Jaguar - 28 stops (c) Seine - 24 stops

(d) Tears of Steel - 17 stops (e) Mercedes - 19 stops (f) Caffe - 23 stops

Fig. 5: HDR video sequences used in the subjective evaluation

The correlation results shown in Tables II and III clearly demonstrate a high to very correlation results between the perceptual/structural QA metrics and the subjective evaluations. Earlier, the evaluation conducted by Mukherjee et al. [1] has shown that the correlation of energy difference metrics (PSNR, logPSNR, Weber MSE) and subjective evaluations are not high.

VI. HDR VIDEO SEQUENCES

This section provides a tone-mapped thumbnail representation of the 39 HDR video sequences used for the objective evaluation. Figures 6 and 7 present tone-mapped thumbnails for each of the sequences along with the average dynamic range.

REFERENCES

[1] R. Mukherjee, K. Debattista, T. Bashford-Rogers, P. Vangorp, R. Mantiuk, M. Bessa, B. Waterfield, and A. Chalmers, "Objective and subjective evaluation of high dynamic range video compression," *Signal Processing: Image Communication*, vol. 47, pp. 426–437, 2016.


(a) s01 - subjective - 21 stops


(b) s02 - subjective - 23 stops


(c) s03 - subjective - 28 stops


(d) s04 - subjective - 24 stops


(e) s05 - subjective - 17 stops


(f) s06 - subjective - 19 stops


(g) s07 - 23 stops


(h) s08 - 23 stops

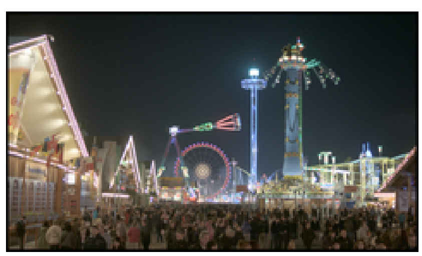

(i) s09 - 23 stops


(j) s10 - 23 stops


(k) s11 - 23 stops

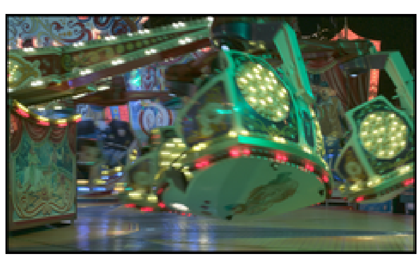

(l) s12 - 23 stops


(m) s13 - 23 stops


(n) s14 - 23 stops

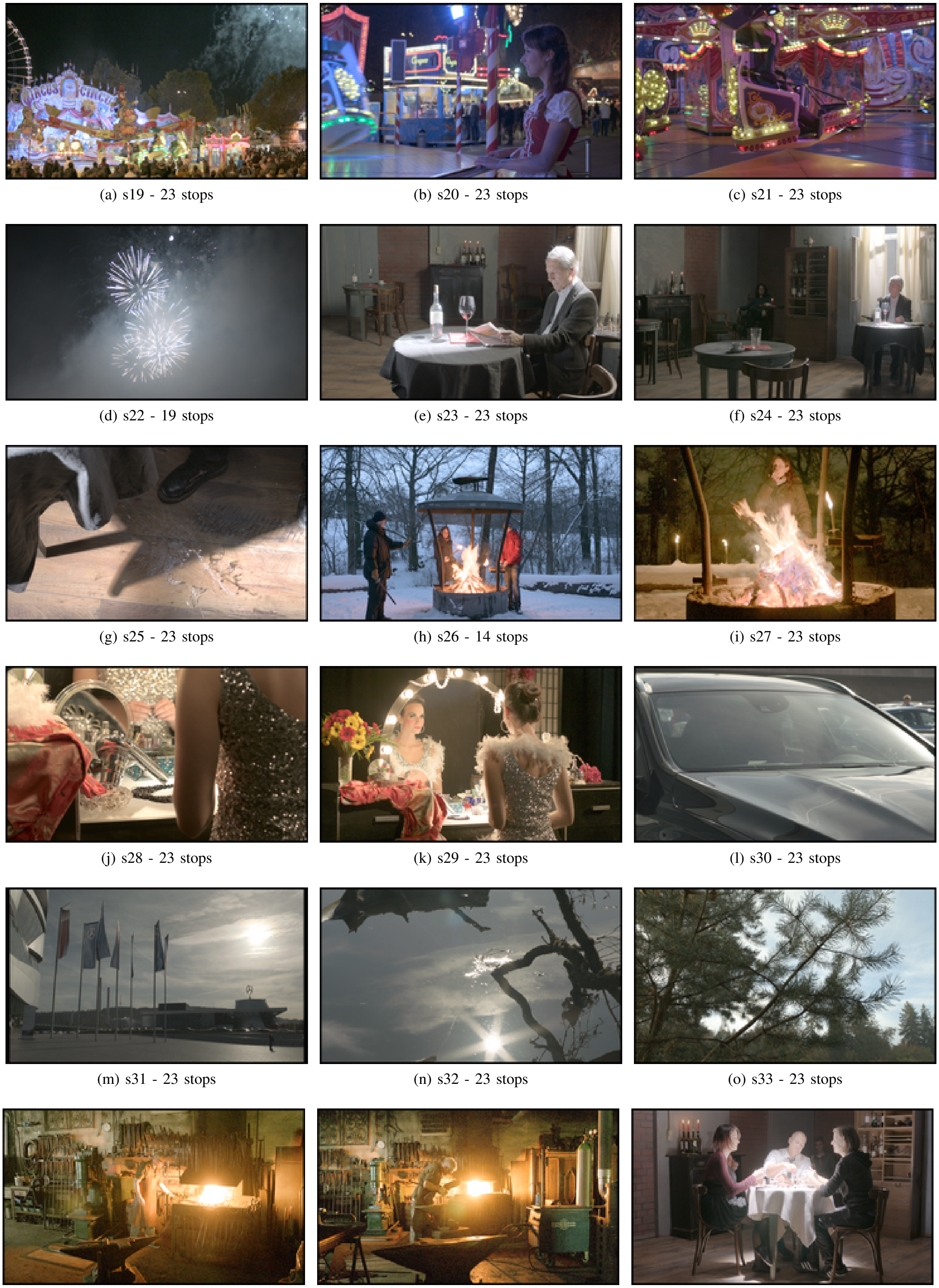

(o) s15 - 23 stops


(p) s16 - 23 stops


(q) s17 - 23 stops

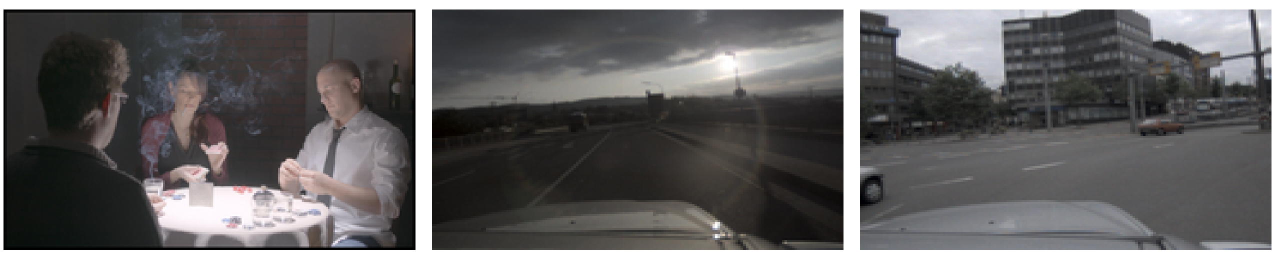

(r) s18 - 23 stops

Fig. 6: HDR video sequences - part I



(a) s19 - 23 stops
(b) s20 - 23 stops
(c) s21 - 23 stops
(d) s22 - 19 stops
(e) s23 - 23 stops
(f) s24 - 23 stops
(g) s25 - 23 stops
(h) s26 - 14 stops
(i) s27 - 23 stops
(j) s28 - 23 stops
(k) s29 - 23 stops
(l) s30 - 23 stops
(m) s31 - 23 stops
(n) s32 - 23 stops
(o) s33 - 23 stops
(p) s34 - 23 stops
(q) s35 - 23 stops
(r) s36 - 23 stops

Fig. 7: HDR video sequences - part II



(a) s37 - 23 stops (b) s38 - 23 stops (c) s39 - 17 stops

Fig. 8: HDR video sequences - part III