Data Adaptive HDR Compression in VVC

Pankaj Topiwala, Madhu Krishnan, Wei Dai {pankajtva at. gmail dot. com}
FastVDO LLC, 3097 Cortona Drive, Melbourne, FL, USA

ABSTRACT

This paper presents an advanced approach to HDR/WCG video coding developed at FastVDO called FVHDR, and built on top of the Versatile Video Coding (VVC) VTM-5.0 test model of the Joint Video Exploration Team, a joint committee of ITU|ISO/IEC. A fully automatic adaptive video process that differs from a known HDR video processing chain (analogous to HDR10, and herein called “anchor”), is used. FVHDR works entirely within the framework of the VVC software model but adds additional tools. These tools can become an integral part of a future video coding standard or be extracted as additional pre- and post-processing chains. Reconstructed video sequences using FVHDR show an improved subjective visual quality to the output of the anchor. Moreover, the resultant SDR content generated by the data adaptive grading process is backward compatible.

Keywords: video compression, high dynamic range, JVET, VTM-5.0, HDR/WCG.

1. INTRODUCTION

HDR/WCG (herein “HDR”) video data have high levels of luminosity, of at least 1000 nits and potentially up to 10000 nits, and with a much wider color gamut, given by the BT. 2020 standard, compared to standard dynamic range (SDR) video content [1]. A standardized approach to coding such content using HEVC has been recently developed called HDR10 and is now being deployed in broadcast and other video applications. The core coding standard utilized in HDR10 is the Main10 Profile of ITU|ISO/IEC HEVC, High Efficiency Video Coding standard. Currently JVET is exploring the coding of HDR content using VVC [2]. As an active member of the committee, FastVDO has developed several effective approaches for HDR video coding that can be implemented either as pre/post processing steps or as normative changes to the core 10-bit encode/decode processes of VVC [2]. We will use an analog of the HDR10 approach as an anchor for comparisons.

2. ANCHOR CODING CHAIN

For our purposes the anchor coding chain is presented in Figure 1. The input video to the encoder is in Y’CbCr 4:2:0 (HDR) format generated from linear light RGB video signal after tone mapping using the ST-2084(PQ OETF) [4] or the HLG transfer function followed by color transformation and then downsampling. The encoding is done with VTM-5.0 profile. The corresponding VTM-5.0 decoder is used to get the reconstructed Y’CbCr 4:2:0 HDR video.

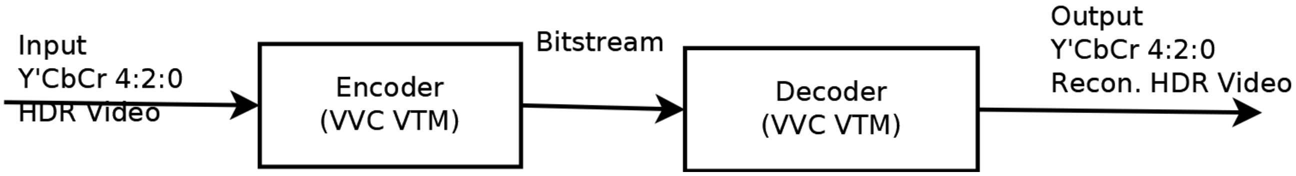


Figure 1: Basic HDR processing chain. The input/output are in a fixed, static transfer function domain (e.g., ST-2084, or HLG). This will serve as our anchor, with VTM -5.0.

3. RECENT WORK IN HDR CODING AT JVET

The Joint Video Exploration Team of the ITU/VCEG|ISO/IEC WG11 (“MPEG”) conducted a Call for Proposals, whose responses arrived in April, 2018, at the San Diego meeting. There were a large number of respondents, many responding to all three areas cited in the Call: SDR, HDR, and 360 video. Among the proposals that had tools specifically targeted for HDR video coding, we note just two: JVET-J0015 (Interdigital/Dolby) [9], and JVET-J0022 (Qualcomm/Technicolor) [10]. J0015 proposed a normative approach to coding HDR video, with an in-loop reshaper function and related tools, while J0022 proposed a pre/post processing approach, with dynamic range adaptation (DRA) applied after normative decoding. These proposals led to various CE elements (of CE12), reviewed at the July, 2018 meeting in Ljubljana. We note that the Breakout Group (BoG) on HDR produced the Table 1. Subsequently, in Jan. 2019 JVET meeting, the in-loop reshaping tool [9][11] was adopted into VTM-4.0. Rather than considering which approach is best, for our purposes, we merely observe that the results of the various tests are mixed, in which some objective metrics may appear to show a gain, and others a loss, while still other data shows modest levels of change that may not be sufficient for adoption. Our main point, observed from several years of working in this field, is that HDR video currently lacks a highly reliable objective metric. Several different metrics are in current testing, which we will not review here, but these can provide conflicting results. Occasionally if they are in sync, they may have some predictive power. But the standard-bearer PSNR fails here, as it becomes dominated by bright areas. Moreover, unlike in traditional SDR video testing, even subjective testing for HDR is non-trivial, with considerations of which monitor, viewing conditions, and scoring regimen.

		RA - Over VTM1.0				
Test#	**Description**	DE100	PSNRL	wPsnrY	wPsnrU	wPsnrV
out-of-loop						
CE12-1	DRA + RefintFlt	10.6%	-6.2%	-4.0%	35.1%	212.0%
CE12-3.1	DRA + RefintFlt + qpHarmoniz	10.3%	-6.6%	-4.4%	34.5%	68.5%
CE12-3.2	DRA + RefintFix + qpHarmoniz	10.4%	-6.6%	-4.4%	34.5%	68.4%
CE12-3.3	DRA + qpHarmoniz	15.6%	-6.5%	-4.3%	70.5%	118.1%
in-loop						
CE12-5	Reshap	11.0%	-2.6%	-2.0%	12.0%	10.3%
CE12-6.1	Reshap + chrQpOffset	-1.3%	-2.6%	-2.0%	0.0%	-0.1%
CE12-6.2	Reshap + chrScaling	2.0%	-2.6%	-2.0%	5.9%	8.9%
CE12-7.1	Reshap + LF_intraOrig_interOrig	-1.1%	-1.7%	-1.3%	0.1%	0.0%
CE12-7.2	Reshap + LF_intraResh_interOrig	-1.4%	-2.4%	-1.8%	-0.1%	-0.1%
CE12-7.3	Reshap + LF_intraResh_interResh	-1.3%	-2.6%	-2.0%	0.0%	-0.1%
CE12-8	Reshap + ROI	-1.2%	-2.2%	-1.7%	0.1%	-0.1%

Table 1 (from JVET-K0552, BoG Report on HDR). This table lists a variety of tools under test. For our purposes, our main observation is that the various objective metrics disagree, providing inconsistent evidence, for algorithms and the metrics.

4. FVHDR SYSTEM DESCRIPTION

In contrast to figure 1, the basic block diagram of the FVHDR approach is represented in Figure 2, which is here depicted as pre/post processing (alternative approaches that are embedded normatively in the codec are also available). A more extensive discussion of FastVDO approaches to HDR coding, in the context of HEVC, are covered in the Technical Report ISO/IEC TR23008-15 [TR2]. These methods are now adopted for VVC. A data adaptive grading (DAG) process is applied to the Y'CbCr 4:2:0 HDR input video. The graded YCbCr/YFbFr 4:2:0 SDR video along with the associated metadata is then encoded with VVC using the equivalent of the Main 10 profile. After decoding the bitstream an inverse grading process is applied to the reconstructed YCbCr/YFbFr 4:2:0 SDR video to generate the Y'CbCr 4:2:0 HDR reconstructed video. This process utilizes the decoded metadata for adaptive degrading. One variant of this method was presented by us at the JVET standards committee, at the Torino, IT, meeting, July, 2017 [8].

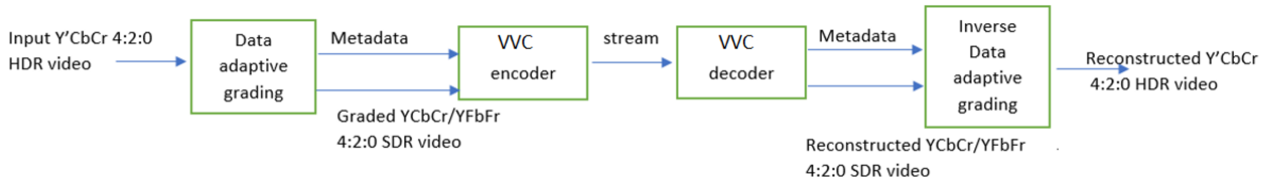


Figure 2: FVHDR coding chain. The difference to figure 1 is that we now use pre/post processing with data-adaptive grading, compared to static transfer functions (such as PQ, HLG).

4.1. DETAILS OF THE DATA ADAPTIVE GRADING (DAG)

The fully automatic DAG process is essentially a point operation on linear light signal. The Y'CbCr 4:2:0 HDR input video is upsampled to Y'CbCr 4:4:4 signal. A color transformation is then done to get a R'G'B' 4:4:4 signal. PQ EOTF is then applied to get linear light RGB$_{HDR}$. The RGB$_{HDR}$ input is then graded using a data adaptive OETF to RGB$_{SDR}$. The RGB$_{SDR}$ content is then mapped to a non-constant luminance (NCL) Y'CbCr or a FastVDO developed YFbFr signal [5]. The resultant YCbCr/YFbFr signal is downsampled to 4:2:0 format using advanced chroma downsampling filters designed by FastVDO [6]. The YCbCr/YFbFr 4:2:0 signal is then encoded using the VTM reference software. Figure 3 describes the sequential steps involved in DAG.

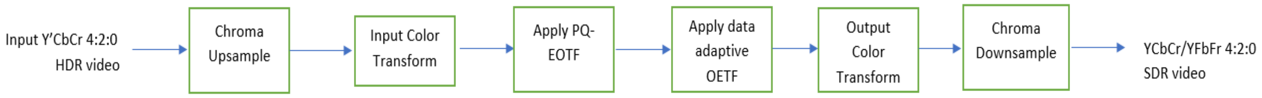


Figure 3: Data Adaptive Grading (DAG) flowchart. A key novel component is the data-adaptive OETF (FVPQ). Note that in this paper we have developed an adaptive version of PQ, though similar ideas can be applied to any transfer function.

Data adaptive OETF: The application of data adaptive OETF involves two steps:

1. Determine $InputPeakBrightness$: The input Y'CbCr 4:2:0 HDR video is analyzed to determine the peak brightness/luminosity of data.
2. Apply a FastVDO modified version of PQ-OETF (herein called **FVPQ-OETF**) to each channel of the input. The FVPQ-OETF function is described as (where Y now stands for individual signal components of RGB$_{HDR}$ and Y_{OETF} for individual signal components of RGB$_{SDR}$):

$$Y_{OETF} = powf(((c2 * powf(Y, m1) + c1)/(1.0 + c3 * powf(Y, m1))), m2) \text{ where}$$

$$m1 = (2610.0 \quad) \,/\, (4096.0 * 4.0)\,,$$

$$m2 = \left(\frac{(2523.0 * 128.0)}{4096}\right) * \left(1 + 0.25 * log\left(\frac{10000}{InputPeakBrightness}\right)\right),$$

$$c1 = (3424.0 \quad) \,/\, 4096.0,$$

$$\text{c}2 = (2413.0 * 32.0) \,/\, 4096.0,$$

$$c3 = (2392.0 * 32.0) \,/\, 4096.0$$

The inverse function (FVPQ-EOTF):

$$Y = powf\left(\frac{dMax\left(0.0, powf\left(Y_{OETF}, \frac{1.0}{m2}\right) - c1\right)}{c2 - c3 * powf\left(Y_{OETF}, \frac{1.0}{m2}\right)}, \frac{1.0}{m1}\right);$$

Here *dMax(a,b)* returns *a* if *a* > *b* else *b*.

Here, $powf(a, b) = a^b$. Thus, the transfer function is adapted based on luminosity. When the parameter equals 10,000, this reduces to the usual PQ function. Figure 4 shows the function for different data luminosity. The entire process described in Figure 3 can be viewed as a single non-linear function. One dimensional (1D) look up tables can be generated to represent the mapping and sent as metadata information. Changes in mean luminosity between frames is used to trigger the transmission of new metadata information.

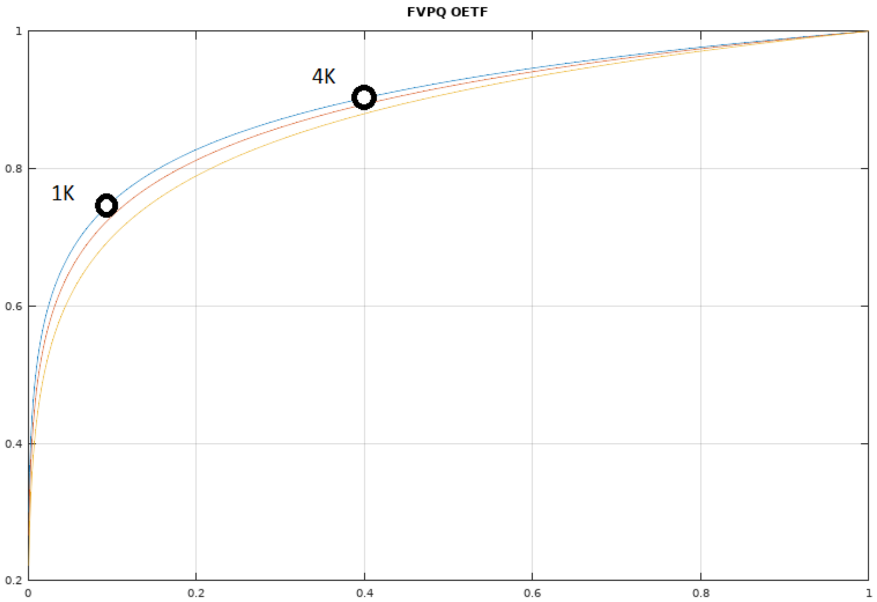


Figure 4: FVPQ OETF, a FastVDO update on the PQ transfer function, for fixed values of InputPeakBrightness = 10000, 4000, and 1000 nits, respectively (at 10K nits it equals the ST-2084 PQ function), in descending order. But when the input data is below 10000 nits, our TF affords more codewords at the brighter end than the standard PQ transfer function. The difference is significant. In fact, a 1K-nit signal using the standard PQ function would only use a portion of the curve (and codeword space); similarly for a 4K-nit signal.

4.2. INVERSE DATA ADAPTIVE GRADING (IDAG)

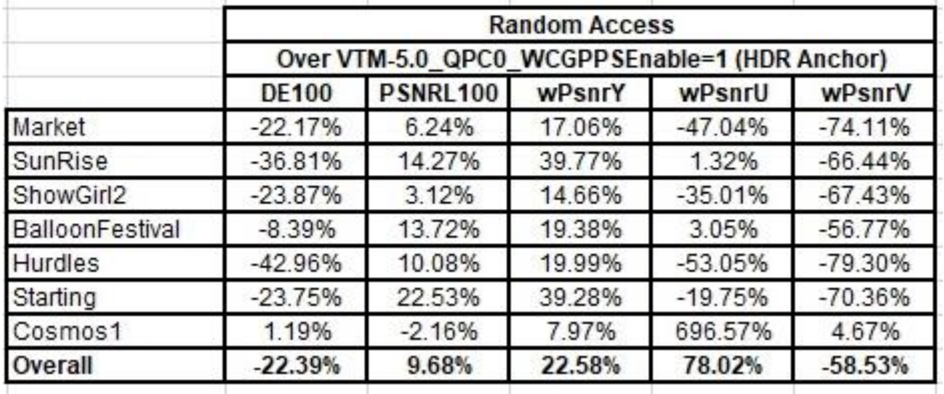
At the decoder side, the bitstream is decoded to get the reconstructed YCbCr/YFbFr 4:2:0 SDR signal and metadata (1D lookup tables). The inverse DAG process uses the lookup tables to map YCbCr/YFbFr 4:2:0 SDR to Y’CbCr 4:2:0 HDR video.

5. EXPERIMENTAL SETUP

The test sequences used in this paper are Y’CbCr 4:2:0 BT.2100 PQ 10-bit sequences. More details on the test sequences are available in [2]. The VVC reference codec VTM-5.0 is used for encoding/decoding process. Bitstreams are generated based on the Main 10 profile under Random access coding constraints. A structural delay of 16 pictures (GOP size of 16) was selected. Intra random access pictures, corresponding to the parameter ‘IntraPeriod’ in the VVC configuration files, are introduced into the bit stream every 32 or 48 pictures for 24 fps, 25 fps and 50 fps sequences, respectively [2]. The codec is run at constant quality with fixed QP values. For each sequence, four rate points are generated corresponding to four different QP values. The simulations were performed in a Rocks cluster distribution with AMD Phenom II X6 1055T’s and Intel i7-6700HQ as the CPU’s and a RAM capacity of 16GB each.

6. RESULTS & CONCLUSION

We apply our methodology to all HDR test sequences by first converting to linear light, and applying our adaptive method, called FVHDR. The objective results obtained by FVHDR with respect to the HDR10-like anchor are presented in tables 2, 3. The modified Bjøntegaard Delta Rate (BDRATE) [7] metric - which is a measure of the integral of the rate difference between two RD curves (FVHDR vs anchor) is used for comparison. A negative BDRATE equates to bitrate savings in percentage. The BDRATE is calculated for tPSNR-Y, DE100, PSNRL100, wPSNR and PSNR metrics respectively. The tPSNR-Y, DE100 and PSNRL100 calculations are done at linear light while wPSNR and PSNR are calculated at the encoder. FVHDR provides a superior visual quality to the output of the anchor processing chain on all test sequences. Figure 5 show the visual quality difference between these two approaches encoded at a similar bitrate (within ±2%). Here we compare cropped sections from a frame of sequence named Market3Clip4000r2 (encoded at 1496 Kbps using the anchor and at 1520 Kbps using FVHDR). The superior visual quality of FVHDR at a similar bitrate is evident from the captured crops. We have thus explained an adaptive HDR coding technology that provides consistently superior performance over baseline anchors. In Table 2 the objective performance of FVHDR is tabulated using the DE100, PSNRL100 and wPSNR metrices. These objective metrics unfortunately are not indicative of visual quality.



	Random Access				
	Over VTM-5.0_QPC0_WCGPPSEnable=1 (HDR Anchor)				
	DE100	PSNRL100	wPsnrY	wPsnrU	wPsnrV
Market	-22.17%	6.24%	17.06%	-47.04%	-74.11%
SunRise	-36.81%	14.27%	39.77%	1.32%	-66.44%
ShowGirl2	-23.87%	3.12%	14.66%	-35.01%	-67.43%
BalloonFestival	-8.39%	13.72%	19.38%	3.05%	-56.77%
Hurdles	-42.96%	10.08%	19.99%	-53.05%	-79.30%
Starting	-23.75%	22.53%	39.28%	-19.75%	-70.36%
Cosmos1	1.19%	-2.16%	7.97%	696.57%	4.67%
Overall	**-22.39%**	**9.68%**	**22.58%**	**78.02%**	**-58.53%**

Table 2. Objective results for FVHDR vs VTM-5.0 for HDR coding.

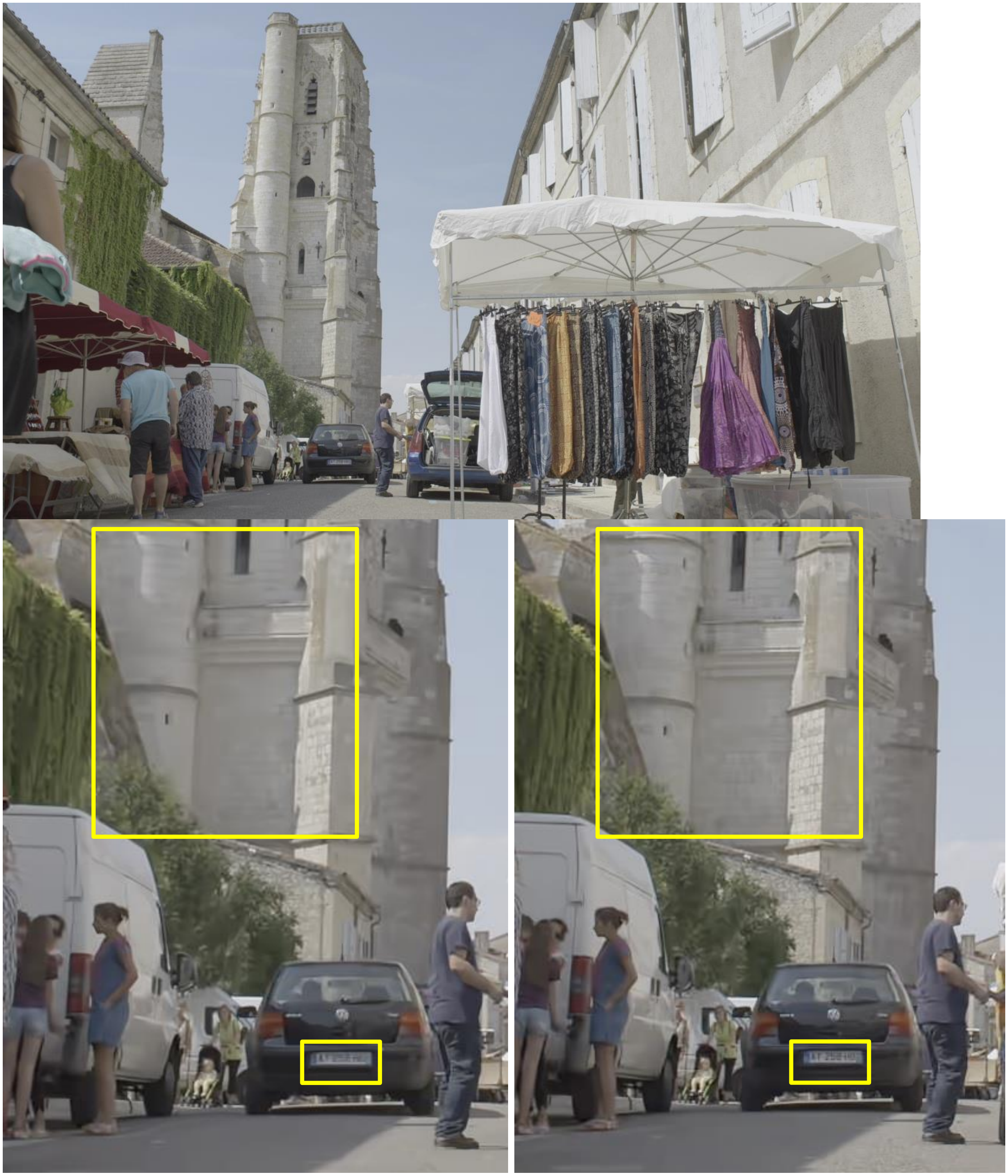


Figure 5. (a) An original frame from the Market3Clip4000r2 HDR test sequence. In the development of VVC, it has been an important test case for HDR coding. (b) Cropped sections from (i) a reconstructed frame from the VVC Anchor (encoded at 1496Kbps) and (ii) a similar reconstructed frame from FVHDR (encoded at 1520Kbps). The FVHDR coding approach appears to retain edges and texture on the church walls, and letters on the registration plate are much more legible.

7. REFERENCES

[1] AHG on HDR and WCG, M33622, ISO/IEC JTC1/SC29/WG11 MPEG, July 2014, Sapporo, Japan.

[2] VTM Test Model. https://jvet.hhi.fraunhofer.de/

[3] JVET-F1001, Algorithm Description of Joint Exploration Test Model 6 (JEM 6), 6th Meeting: Hobart, AU, 31 March – 7 April 2017.

[4] Society of Motion Picture and Television Engineers ST 2084 (2014), High Luminance EOTF.

[5] Pankaj Topiwala and Chengjie Tu, "New Invertible Integer Color Transforms Based on Lifting Steps and Coding of 4:4:4 Video", JVT-I015r7, San Diego, Sept., 2003.

[6] Trac D. Tran, Lijie Liu, and Pankaj Topiwala, "Advanced Extended Spatial Re-sampling Filters For SVC", JVT-V030, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 22nd meeting, Marrakech, Morocco, Jan. 2007.

[7] G. Bjøntegaard, "Calculation of average psnr differences between rd-curves," VCEG-M33, 13th VCEG meeting, Austin, Texas, March 2001.

[8] P. Topiwala, W. Dai, M. Krishnan, "FastVDO Response to the JVET CfE for HDR", JVET-G0021, Torino, IT, July, 2017.

[TR2] ISO/IEC TR23008-15, "Signalling, Backward Compatibility and Display Adaptation for HDR/WCG Video," 2018.

[9] X. Xiu et al, "Description of the SDR, HDR, and 360 video coding technology proposal by InterDigital Communications and Dolby Laboratories," JVET-J0015, San Diego, CA, April, 2018.

[10] P. Bordes et al, "Description of the SDR, HDR, and 360 video coding technology proposal by Qualcomm and Technicolor – medium complexity version," JVET-J0022, San Diego, CA, April, 2018.

[11] T. Lu, F. Pu, P. Yin, W. Husak, S. McCarthy, T. Chen, "CE12: Mapping functions (test CE12-1 and CE12-2)," Document of Joint Video Experts Team (JVET), JVET-M0427, Marrakech, MA, Jan 2019.