This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Analysis of the perceptual quality performance of different HEVC coding tools

JAVIER RUIZ ATENCIA[1], OTONIEL LÓPEZ GRANADO[1], MANUEL PÉREZ MALUMBRES[1], (MEMBER, IEEE), MIGUEL ONOFRE MARTÍNEZ-RACH[1], (MEMBER, IEEE), AND GLENN VAN WALLENDAEL[2].

[1]Computer Engineering Department at Miguel Hernández University of Elche (e-mail: javier.ruiza, otoniel, mels, mmrach@umh.es)
[2]IDMedia-Lab, Ghent University, Belgium (e-mail: Glenn.VanWallendael@UGent.be)

Corresponding author: Javier Ruiz Atencia (e-mail: javier.ruiza@umh.es).

This research was supported by the Spanish Ministry of Science, Innovation and Universities under Grant RTI2018-098156-B-C54, co-financed by FEDER funds (MINECO/FEDER/UE) and by the Ministry of Innovation, Universities, Science and Digital Society of the Valencian Government under Grant GV/2019/020.

ABSTRACT Each new video encoding standard includes encoding techniques that aim to improve the performance and quality of the previous standards. During the development of these techniques, PSNR was used as the main distortion metric. However, the PSNR metric does not consider the subjectivity of the human visual system, so that the performance of some coding tools is questionable from the perceptual point of view. To further explore this point, we have developed a detailed study about the perceptual sensibility of different HEVC video coding tools. In order to perform this study, we used some popular objective quality assessment metrics to measure the perceptual response of every single coding tool. The conclusion of this work will help to determine the set of HEVC coding tools that provides, in general, the best perceptual response.

INDEX TERMS HEVC, perceptual coding, transform skip, RDOQ, deblocking filter, SAO, CSF, perceptual metrics

I. INTRODUCTION

HIGH Efficiency Video Coding (HEVC) is the latest video coding standard in force developed by the Joint Collaborative Team on Video Coding (JCT-VC) of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) standardization organizations. During the development of the standard, a set of working draft specifications, including the accepted proposals, was published. In addition, the HEVC Test Model (HM) reference software was provided, so that the different coding techniques proposed could be tested.

In 2013, the first version of the standard was released. New versions of the standard and the reference software, including the multi-view extensions (MV-HEVC), the range extensions

(RExt), and the scalability extensions (SHVC), have been launched since then.

The main goal of the HEVC standard was to reduce the bit rate by up to 50% while maintaining the same subjective quality as the previous H.264/AVC standard, without increasing the complexity of the encoder. To accomplish this goal, the HEVC standard incorporates numerous coding techniques that attempt to reduce the bit rate without increasing the distortion. Many of these techniques are based on the previous H.264/AVC standard, while other novel features, such as Quad-tree partitioning or the Sample Adaptive Offset (SAO) filter, were also included.

Some coding tools in the HEVC standard include approaches that deal with the non-linear behavior of the Human Visual System (HVS), in order to take into account the subjective quality perceived by humans during the encoding process. In particular, HEVC provides the SCaling List (SCL) coding tool, which applies a non-uniform quantization to the transformed coefficients, depending on the HVS contrast sensitivity associated to their frequencies. The main idea is that higher quantization can be applied to the areas of the scene for which the HVS is less sensitive, i.e., the Just Noticeable Distortion (JND) concept [1]. Some studies also include HEVC profiles to manage the luminance masking effect for High Dynamic Range (HDR) video sequences, as in [2] [3], where the authors apply a non-uniform quantization profile based on the Intensity Dependent Quantization [4]; it is adaptively applied to each frame based on a tone-mapping operator. An important bit rate reduction is obtained for the same quality that was measured with the specifically designed HDR-VPD-2 quality assessment metric [5] and also through subjective tests.

Although the coding techniques incorporated in the HEVC standard have proven to be capable of reducing the bit rate, it cannot be guaranteed that they are optimized from a perceptual point of view. During the development of the coding techniques, the Peak Signal-to-Noise Ratio (PSNR) metric was used to measure the distortion. PSNR, like Mean Square Error (MSE), provides a quality score based on the pixel differences between the original and reconstructed images. It is well known that these metrics do not accurately reflect the perceptual assessment of quality [6] [7] [8] [9]. However, in the existing literature, there seems to be conflicting evidence about the accuracy of PSNR as a video quality metric. In [10], the authors proved that PSNR follows a monotonic relationship with subjective quality in the case of full frame rate encoding, when the video content and the video encoder are fixed.

So, in order to properly asses the perceptual (i.e., HVS-like) performance of HEVC coding tools, we need to employ quality assessment metrics that provide quality scores highly correlated with the quality perceived by humans. By doing this, we will ensure that the HEVC coding tools are always evaluated to maximize the perceptual performance of the overall encoder, avoiding, as much as possible, the deployment of cumbersome and time-consuming subjective tests.

In this study, we analyze the perceptual performance of the several coding tools of the HEVC Test Model (HM) software, which concerns the visual quality of a reconstructed video sequence. We have encoded the whole set of video sequences included in the HEVC common test conditions [11] with different configuration setups in order to analyze their perceptual response, trying to understand which encoder configuration maximizes the averaged perceptual quality of the reconstructed video sequences. So, on the one hand, we modify the encoder by changing its configuration, and on the other hand, we use multiple sequences (different content) to obtain this average. Therefore, under these conditions, PSNR should not be used as a reference metric to obtain perceptually based conclusions [12].

Each configuration setup will determine which coding tools are enabled, so we may analyze not only their individual contribution to the perceptual performance but also their contribution in combination with other coding tools. In order to measure the video quality, we use a set of well-known image objective quality assessment metrics, as well as the new Video Multimethod Assessment Fusion (VMAF) quality metric developed by Netflix [13].

The main contribution of this paper is based on the R/D performance analysis of several HEVC coding tools, in order to properly assess their impact on the perceptual quality of the decoded video. Most of the available studies about HEVC coding tools in the literature only work with the PSNR, but very few of them are interested in perceptual behavior; usually, they focus only on a specific part of the encoder. We have exhaustively analyzed the impact of different coding tools on the perceptual quality of the reconstructed videos, showing results that differ from the results provided by PSNR. These results may be useful for future studies in order to configure the video encoder to maximize the perceptual R/D performance.

The rest of the article is structured as follows. An overview of the different HEVC coding tools under study is presented in Section II. In Section III, the methodology used in this study is explained. Section IV shows the experimental results, while in Section V, we provide a brief discussion of the obtained results. Finally, Section VI summarizes the conclusions of this study, and some future research lines are pointed out.

II. HEVC PARAMETERS

The HEVC standard includes many configuration parameters that are used to enable or disable coding tools that improve the reconstructed quality, reduce bit stream size, or simplify encoder complexity.

These parameters allow us to tune the coding structure, motion estimation, quantization, entropy coding, slice coding, deblocking filter, and rate control, among others [14]. Inside these main coding parameter blocks, the user can enable or disable the use of any parameter, as well as create a user-defined behavior for a given parameter. For example, in the deblocking filter parameter block, the user can enable or disable the loop filter and also define the use of the loop filter across the slice boundaries (subparameter: LFCrossSliceBoundaryFlag). However, some of these user-defined behaviors for some coding tools (subparameters) may affect the subjective quality. In this paper, we will focus on the general behavior of the HEVC codec when enabling or disabling the main coding parameter inside each parameter block.

In this work, we have selected the following configuration parameters for evaluation, since they have a high impact on the visual quality of the decoded video sequence: *Scaling List, Deblocking Filter, SAO Filter, Rate-Distortion Optimized Quantization*, and *Transform Skip*.

A. SCALING LIST

The HVS is not able to detect all spatial frequencies with the same accuracy [15]. Numerous studies over the past few decades have characterized the Contrast Sensitivity Function (CSF) [16] [17] [18] as the response of our HVS to contrast variations, showing that the human eye is least sensitive to the highest and lowest frequencies.

This CSF is implemented in the quantification stage of the HEVC standard and can be modified by the Scaling List parameter, with three available options. By default, the encoder applies a constant quantizer step size for all transform coefficients (ScalingList = 0) that does not consider the subjectivity of the HVS. However, the HEVC standard includes pre-defined weighting matrices (ScalingList = 1) that incorporate an implementation of the CSF. These non-flat matrices (like the one shown at Figure 2-b) define an additional scaling of the quantizer step, which varies with the transformed coefficient position, i.e., the base function frequency [19].

The results of the study carried out by [20] showed that, on average, the use of the weight

$$\begin{bmatrix} 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16\\ 16&16&16&16&16&16&16&16 \end{bmatrix}$$

(a) 8×8 flat quantization matrix

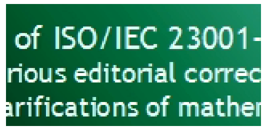
$$\begin{bmatrix} 16&16&16&16&17&18&21&24\\ 16&16&16&16&17&19&22&25\\ 16&16&17&18&20&22&25&29\\ 16&16&18&21&24&27&31&36\\ 17&17&20&24&30&35&41&47\\ 18&19&22&27&35&44&54&65\\ 21&22&25&31&41&54&70&88\\ 24&25&29&36&47&65&88&115 \end{bmatrix}$$

(c) 8×8 weight quantization matrix

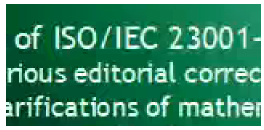
FIGURE 1. Default 8×8 quantization matrices for (a) ScalingList = 0 and (b) ScalingList = 1.

(a) (b)

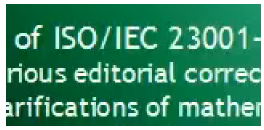
FIGURE 2. Example of using deblocking filter in BlowingBubbles frame, encoded at QP37: (a) DB disabled, (b) DB enabled.



(a)



(b)



(c)

FIGURE 3. Example of using SAO filter in SlideEditing frame, encoded at QP32: (a) Original uncompressed (b) SAO disabled, (c) SAO enabled.

quantization matrices provides better subjective quality results.

B. DEBLOCKING FILTER

This filter reduces the effect of blocking artifacts that are inherent in the nature of the encoder. It is used after block reconstruction, but its implementation is done within the coding loop, i.e., the reconstructed and filtered blocks will be taken as reference for other blocks (in-loop filter).

Its implementation is similar to the one used in the H.264 standard [21], but it is somewhat more simplified. In HEVC, the decoder can adaptively choose between applying two levels of the deblocking filter (normal or strong) or not applying it, depending on the adjacent blocks and a certain threshold [22]. As an example, Figure 2 compares the use of the deblocking filter on a part of a decoded picture. As can be seen, the grid effect disappears when the filter is active; however, regions that should not be filtered out are also blurred.

In [22], the authors argue that applying the deblocking filter increases the objective and subjective quality of the decoded video sequences.

C. SAO FILTER

The Sample Adaptive Offset (SAO) filter is a new algorithm integrated in the HEVC standard. It is located after the deblocking filter, and together, they form the so-called in-loop filter stage.

The main purpose of the SAO filter is to reduce distortion in the samples. To this end, the samples are classified into different categories, obtaining an offset for each of them. There are two sample processing techniques, band offset and edge offset. The algorithm will adaptively decide on the best strategy to use. The offset value is transmitted through the bit stream, while the classification of the samples is performed on both the encoder and decoder sides to reduce the information to be transmitted [23].

In [23], authors explain that using the SAO filter can provide about coding gains of 3.5% on average. To measure this gain, they have used the Bjøntegaard-Delta Rate (BD-Rate) metric [24], which uses the PSNR, a non-subjective

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

14	3	7	**1**
9	-4	1	0
2	-4	0	0
-1	0	1	0

(a)

14	3	7	**0**
9	-4	1	0
2	-4	0	0
-1	0	1	0

(b)

FIGURE 4. Modified coefficient value example in 4×4 transform sub-block example for sign data hiding.

metric. Regarding the subjective quality, the authors state that, based on experiments carried out by themselves, an improvement in quality is generally perceived. This improvement is higher in synthetic images, as shown in Figure 3, where SAO significantly improves the visual quality by suppressing the ringing artifacts near edges.

D. RATE-DISTORTION OPTIMIZED QUANTIZATION (RDOQ)

In the video encoder, optimizing the quantification process has a significant impact on the compression efficiency. The HEVC standard does not specify the quantization function, giving the encoder some flexibility in implementing it. HEVC includes, since version 13 of the reference software, a more sophisticated implementation of the quantization scheme called rate-distortion optimization quantization (RDOQ).

The purpose of RDOQ is to find the optimal or suboptimal set of quantized transform coefficients representing residual data in an encoded block. RDOQ calculates the image distortion (introduced by the quantization of transformed coefficients) in the encoded block and the number of bits needed to encode the corresponding quantized transform coefficients. Based on these two values, the encoder chooses, among different coefficient blocks, the block which provides the better Rate Distortion (RD) cost [25].

Note that RDOQ is an effective method in terms of increasing the R/D performance. However, in [26], the authors claim that the PSNR-based mathematical reconstruction quality improvement attained by this technique is perceptually negligible in terms of how the human observer interprets the perceived quality of the compressed video data.

E. TRANSFORM SKIP

The Transform Skip parameter in the HEVC standard allows the encoder to bypass the transformation stage. In this way, the prediction errors are coded directly in the spatial domain.

During the development of HEVC, three transform skip modes were proposed and tested, but the standardization committee finally decided to use a single mode, the skipping transform in both the vertical and horizontal directions [27]. This mode was found to improve the compression of synthetic video sequences such as remote desktop, slideshows, etc.

Finally, the HM reference software includes an additional parameter, called TransformSkipFast, which enables or disables reduced testing of the transform-skipping mode decision in order to speed it up.

F. SIGN DATA HIDING

The transform coefficient coding in HEVC includes an option, called sign data hiding or sign bit hiding, that hides the coding of the sign flag of the first non-zero coefficient at the parity of the absolute sum of the coefficients. If the parity does not match the sign of the first non-zero coefficient and there are a sufficient number of significant coefficients, the encoder will modify the amplitude of a block coefficient until the desired sign is obtained [28].

As an example, in Figure 4-a, we have a transformed sub-block of size 4 x 4, whose absolute sum is 47. By convention, an even value would derive a negative sign for the first non-zero coefficient. Following the block in zig-zag order from left to right and from top to bottom, the first non-zero coefficient is 14, which has a positive sign. This is why the encoder changes the value of a coefficient (Figure 4-b) so that the absolute sum will be even. The selection of the coefficient to be modified is determined by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

the Rate-Distortion criteria, which chooses the coefficient with the lowest R/D cost.

This compression technique achieves an average BD-Rate reduction of 0.6% for the All Intra coding mode. Modifying the coefficient values tends to increase the distortion, so the BD-Rate gain is obtained thanks to the rate reduction provided by this technique and not due to a quality increase.

III. METHODS AND PROCEDURES

In this work we have followed the indications established by the Common Test Condition [11]. This document defines a regulatory framework establishing a set of defined sequences and several base configurations for HM. The set of test sequences are classified in six large groups (A–F). The classes A, B, C, and D represent video sequences with different contents, video resolutions, frame rates, and bit depths. Class E is focused on head and shoulders videos typically used in video conference applications, and class F is devoted to computer generated videos and content screen applications (no natural video sequences).

In this work, we have focused only on the All Intra Main (AI Main) configuration mode, and therefore, no temporal processing and analysis was performed. We have only used the All Intra coding mode under the following criteria: (a) As most perceptual metrics are only available for images (not videos), the objective video quality measurement would be more accurate when using the All Intra coding mode since these metrics are unable to capture the motion-related artifacts. (b) Most of the coding tools analyzed in this paper have a direct impact on the reconstruction quality as a result of a prediction process where the residual error is quantized and the entropy is encoded. So, with the independence of using spatial or temporal prediction, the quality distortion is due to the quantization of the prediction error. If two prediction blocks (one spatial and the other temporal) produce the same residual error, the reconstruction quality should be the same for a given quantization value, so the quality of the prediction and not the prediction itself (temporal or spatial) determines the final reconstruction quality.

TABLE I. HEVC test sequences

Class	Sequence name	Resolution	Frame count	Frame rate	Bit depth
A	Traffic	2560x1600	150	30	8
	PeopleOnStreet		150	30	8
	Nebuta		300	60	10
	SteamLocomotive		300	60	10
B	Kimono	1920x1080	240	24	8
	ParkScene		240	24	8
	Cactus		500	50	8
	BQTerrace		600	60	8
	BasketballDrive		500	50	8
C	RaceHorses	832x480	300	30	8
	BQMall		600	60	8
	PartyScene		500	50	8
	BasketballDrill		500	50	8
D	RaceHorses	416x240	300	30	8
	BQSquare		600	60	8
	BlowingBubbles		500	50	8
	BasketballPass		500	50	8
E	FourPeople	1280x720	600	60	8
	Johnny		600	60	8
	KristenAndSara		600	60	8
F	BaskeballDrillText	832x480	500	50	8
	ChinaSpeed	1024x768	500	30	8
	SlideEditing	1280x720	300	30	8
	SlideShow	1280x720	500	20	8

Table I defines the set of test sequences used in this paper.

In order to analyze the R/D performance of the coding tools described in the previous section, we use the Bjontegaard BD-Rate metric [24], which shows the rate-distortion performance. We followed the instructions defined in the HEVC conformance test standard [11]; the QPs 22, 27, 32, and 37 are used to conform the PSNR curves that allow the computation of the BD-Rate performance. As we are using other objective perceptual metrics like MS-SSIM and VMAF, we decided to add one more QP (QP = 42) in order to better fit the dynamic range response of these metrics, and as a consequence, provide more accurate BD-Rate results.

The BD-Rate values have been obtained for each metric and each coding configuration. These values show the bit rate savings (in percentage) between two rate-distortion curves. Due to the fact that the BD-Rate calculation was initially developed for the PSNR metric using third degree polynomial interpolation and four values per curve, the use of this algorithm applied to other metrics and with five points (QP values) per curve is not always optimal. Therefore, the interpolation method has been replaced by the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [29] for higher accuracy.

To evaluate the influence of each parameter on the perceptual quality described in Section II, all test sequences have been coded by switching these parameters on and off, resulting in 64 configuration setups. These setups have been run using the reference software HEVC Test Model (HM) version 16.20 [30].

Due to the large number of measurements to be made, the use of subjective tests such as DMOS has been ruled out. Instead, we have proposed obtaining numerical values from Bjøntegaard-Delta rate measurements using the following objective quality metrics: SSIM, MS-SSIM, VIF, PSNR-HVS-M, and VMAF.

The SSIM (Structural Similarity) [9] and the MS-SSIM (Multi-Scale SSIM) [8] metrics are based on the hypothesis that the HVS is highly adapted to extract structural information from the scenes. Both metrics consider luminance, contrast, and structure information of the scenes, whereas MS-SSIM also considers the scale.

The VIF (Visual Information Fidelity) [31] metric uses the Natural Scene Statistics (NSS) model along with an image degradation model and components of the HVS to obtain the quality information.

The PSNR-HVS-M metric [32], a modified version of the PSNR, considers the contrast sensitivity function (CSF) and the between-coefficient contrast masking of DCT basis functions.

These metrics, unlike the PSNR, attempt to characterize the subjectivity of the HVS and do not include temporal information in their quality assessment algorithms.

The newest perceptual quality metric is the VMAF metric, developed by Netflix [13]. Unlike the previous metrics, VMAF makes use of novel machine learning techniques to estimate the result that would be obtained through subjective tests. To do that, this metric has been trained with inputs from real DMOS tests as well as three algorithms: VIF, DLM (Detail Loss Measure) [33], and TI (Temporal Perceptual Information) [34]. VIF measures the information fidelity loss, while DLM and TI measure the detail loss and the amount of motion, respectively.

Other works in the literature use VMAF: (a) in [35], the authors proved a strong correlation between subjective DMOS studies and the VMAF values obtained for a set of 4K sequences, (b) in [36], the authors also show a high correlation with the MOS values obtained for HD and UHD content, and (c) in [37], an analysis of different quality metrics for multi-resolution adaptive streaming showed that VMAF obtained the highest correlation with perceptual quality.

We have to say that the VMAF metric can be used for just one frame or for the whole video sequence. As the rest of the quality metrics only work at frame level, we decided to use VMAF for each frame in order to be coherent with the experiment setup and to avoid undesired effects when comparing all quality metrics results.

Regarding rate-distortion curves obtained in this work, the reference rate-distortion curve was obtained with the default All Intra Main configuration, whose parameters are shown in Table II.

IV. EXPERIMENTAL RESULTS

In this section, we show the results obtained after coding the set of test sequences when enabling and disabling the coding tools described in Section II with respect to the default configuration.

In order to measure the R/D performance of the different HEVC coding tool configurations, we have used the BD-Rate metric, as described in the previous section. The results from Tables III-A to III-F are provided by the BD-Rate metric, showing the R/D performance of the different coding tool setups measured by different perceptual quality metrics for all video sequences under evaluation (classes A to F). Negative values in these tables correspond

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

TABLE II. Default values of the analyzed parameters

Parameter	Value
QP	22, 27, 32, 37, 42
ScalingList	0
LoopFilterDisable	0
SAO	1
RDOQ	1
RDOQTS	1
TransformSkip	1
TransformSkipFast	1
SignHideFlag	1

to BD-Rate reductions or perceptual gains, and positive values correspond to BD-Rate increases or perceptual losses, with respect to the default or reference values (first row in Tables III-A to III-F).

In these tables, we have omitted the SignHideFlag coding tool, as it does not provide significant changes in the image distortion (see Section IV-F). The SignHideFlag parameter is enabled in all tests, since this is its default value.

The complete set of tables, including the SignHideFlag parameter, for each of the classes, as well as for each specific video sequence, are available at the GATCOM research group's website [38].

To make the data in the tables easier to read, we have highlighted the cells with BD-Rate reductions as a heat map. The higher the reduction is, the greener the cell is. Each row corresponds to a specific permutation of the configuration parameters; the first row, highlighted in bold, shows the reference setup (default settings).

The first column is the permutation number, and it is used only as a reference in the text. The next five columns correspond to the enabling/disabling status of the following coding tools: SCL corresponds to ScalingList; SAO corresponds to SAO filter; DB corresponds to the inverse logic of the LoopFilterDisable parameter, that is, disabling DB means disabling the Deblocking filter; RDOQ includes both the RDOQ and RDOQTS coding tools; and TrSk includes both TransformSkip and TransformSkipFast.

The values scored by the metrics are not normalized, so each metric provides results in a different scale. However, we express the results in terms of the BD-Rate performance metric (percentage of rate reduction/increase), so we can compare results, hiding the real scale of each metric.

We have also performed a time profile of every single coding tool analyzed in this study to determine their average coding complexity. Considering both evaluation metrics, the perceptual R/D and the coding complexity, we may propose the proper coding tool configuration that better perceptual results provide with a balanced coding complexity.

In the following subsections, we will describe the results obtained showing the perceptual behavior of each coding tool under study, and in the next section, an analysis and discussion of these results will be provided.

A. SCALING LIST

When activating the ScalingList coding tool, a non-uniform quantization based on the contrast sensitivity function (CSF) is applied in the encoder quantization process. By default, it is disabled, so we have analyzed the perceptual influence of enabling it.

The results show that enabling the Scaling List parameter is perceptually beneficial for almost all settings and perceptual metrics. This can be seen by comparing rows 1 to 16 (SCL disabled) with rows 17 to 32 (SCL enabled) of the result tables, where the second group of coding settings generally has a lower BD-Rate value than the first group. Taking into account the base configuration (row 1), just by enabling only the ScalingList coding tool, all perceptual metrics report BD-Rate reductions for all test video sequence classes. Regarding SCL coding complexity, when it is enabled the average encoding time increases between 3.47% and 8.44%, depending on the applied quantization (QP), as shown in Table IV.

From the results in Tables III-A to III-F, we can extract the following main results: (a) when

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

TABLE III-A. Average coding performance [% BD-Rate] for Class A

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
1	**0**	**1**	**1**	**1**	**1**	0	0	0	0	0
2	0	1	1	1	0	−0.03	−0.03	0.06	−0.04	−0.04
3	0	1	1	0	1	5.6	2.29	11.93	3.84	5.38
4	0	1	1	0	0	5.54	2.24	11.95	3.8	5.34
5	0	1	0	1	1	0.26	1.03	1.32	−0.27	1.51
6	0	1	0	1	0	0.24	1	1.38	−0.31	1.46
7	0	1	0	0	1	5.91	3.46	13.48	3.61	7.05
8	0	1	0	0	0	5.86	3.42	13.51	3.57	7
9	0	0	1	1	1	0.18	0.21	−2.08	0.03	0.26
10	0	0	1	1	0	0.15	0.19	−2.03	−0.01	0.21
11	0	0	1	0	1	5.78	2.47	10.02	3.82	5.61
12	0	0	1	0	0	5.73	2.43	10.03	3.8	5.56
13	0	0	0	1	1	0.38	1.56	−3.32	−0.47	3.05
14	0	0	0	1	0	0.35	1.53	−3.27	−0.51	2.99
15	0	0	0	0	1	6.01	4.08	8.68	3.28	8.65
16	0	0	0	0	0	5.95	4.03	8.69	3.24	8.59
17	1	1	1	1	1	−0.38	−0.2	−0.18	−0.34	−0.63
18	1	1	1	1	0	−0.44	−0.24	−0.13	−0.38	−0.68
19	1	1	1	0	1	4.28	1.41	10.71	2.43	3.41
20	1	1	1	0	0	4.23	1.38	10.76	2.4	3.37
21	1	1	0	1	1	−0.11	0.81	1.12	−0.62	0.85
22	1	1	0	1	0	−0.17	0.77	1.18	−0.66	0.79
23	1	1	0	0	1	4.59	2.56	12.24	2.18	5.01
24	1	1	0	0	0	4.54	2.53	12.29	2.15	4.96
25	1	0	1	1	1	−0.19	0	−2.35	−0.34	−0.38
26	1	0	1	1	0	−0.25	−0.04	−2.29	−0.38	−0.43
27	1	0	1	0	1	4.46	1.6	8.78	2.42	3.63
28	1	0	1	0	0	4.4	1.57	8.86	2.39	3.59
29	1	0	0	1	1	0	1.34	−3.58	−0.85	2.37
30	1	0	0	1	0	−0.06	1.29	−3.53	−0.89	2.31
31	1	0	0	0	1	4.68	3.16	7.45	1.86	6.56
32	1	0	0	0	0	4.63	3.12	7.52	1.83	6.51

TABLE III-B. Average coding performance [% BD-Rate] for Class B

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
1	**0**	**1**	**1**	**1**	**1**	0	0	0	0	0
2	0	1	1	1	0	−0.07	−0.06	−0.03	−0.06	−0.08
3	0	1	1	0	1	5.27	3.74	12.66	5.59	7.18
4	0	1	1	0	0	5.22	3.7	12.69	5.56	7.13
5	0	1	0	1	1	0.33	1.4	1.24	0.08	2.12
6	0	1	0	1	0	0.25	1.34	1.22	0.01	2.03
7	0	1	0	0	1	5.67	5.35	14.06	5.71	9.52
8	0	1	0	0	0	5.62	5.31	14.08	5.68	9.47
9	0	0	1	1	1	0.4	0.49	−1.92	−0.06	0.45
10	0	0	1	1	0	0.32	0.43	−1.95	−0.13	0.37
11	0	0	1	0	1	5.68	4.22	10.34	5.36	7.62
12	0	0	1	0	0	5.63	4.18	10.34	5.33	7.56
13	0	0	0	1	1	0.48	2.3	−2.94	−0.2	4.12
14	0	0	0	1	0	0.4	2.23	−2.98	−0.27	4.03
15	0	0	0	0	1	5.78	6.38	9.23	5.23	11.68
16	0	0	0	0	0	5.73	6.34	9.23	5.19	11.62
17	1	1	1	1	1	−0.72	−0.48	−0.47	−0.6	−1.06
18	1	1	1	1	0	−0.78	−0.52	−0.52	−0.71	−1.17
19	1	1	1	0	1	2.76	1.91	9.85	2.53	3.14
20	1	1	1	0	0	2.71	1.87	9.85	2.45	3.06
21	1	1	0	1	1	−0.4	0.9	0.75	−0.57	0.95
22	1	1	0	1	0	−0.45	0.85	0.71	−0.68	0.83
23	1	1	0	0	1	3.16	3.45	11.26	2.58	5.22
24	1	1	0	0	0	3.11	3.41	11.25	2.51	5.13
25	1	0	1	1	1	−0.34	−0.01	−2.47	−0.7	−0.63
26	1	0	1	1	0	−0.39	−0.05	−2.52	−0.82	−0.74
27	1	0	1	0	1	3.17	2.37	7.59	2.33	3.56
28	1	0	1	0	0	3.11	2.34	7.57	2.24	3.46
29	1	0	0	1	1	−0.26	1.77	−3.5	−0.89	2.92
30	1	0	0	1	0	−0.31	1.73	−3.55	−1.01	2.8
31	1	0	0	0	1	3.27	4.46	6.5	2.11	7.32
32	1	0	0	0	0	3.22	4.41	6.47	2.02	7.22

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

TABLE III-C. Average coding performance [% BD-ate] for Class C

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
1	**0**	**1**	**1**	**1**	**1**	0	0	0	0	0
2	0	1	1	1	0	−0.23	−0.22	−0.16	−0.18	−0.26
3	0	1	1	0	1	2.1	2.89	10.43	4.54	5
4	0	1	1	0	0	1.98	2.76	10.34	4.48	4.81
5	0	1	0	1	1	0.79	1	0.85	0.38	1.65
6	0	1	0	1	0	0.57	0.78	0.67	0.2	1.38
7	0	1	0	0	1	3.01	4.06	11.25	4.95	6.75
8	0	1	0	0	0	2.89	3.92	11.14	4.88	6.55
9	0	0	1	1	1	0.71	0.59	−1.84	0.25	0.65
10	0	0	1	1	0	0.5	0.38	−2.06	0.07	0.4
11	0	0	1	0	1	2.89	3.56	8.02	4.78	5.76
12	0	0	1	0	0	2.79	3.44	7.85	4.72	5.58
13	0	0	0	1	1	1.39	2.02	−2.71	0.72	3.77
14	0	0	0	1	0	1.17	1.8	−2.93	0.54	3.51
15	0	0	0	0	1	3.73	5.33	7.13	5.32	9.14
16	0	0	0	0	0	3.63	5.19	6.96	5.25	8.95
17	1	1	1	1	1	−0.02	−0.2	−0.32	−0.09	−0.22
18	1	1	1	1	0	−0.27	−0.44	−0.55	−0.34	−0.56
19	1	1	1	0	1	0.8	1.22	7.75	2.99	2.97
20	1	1	1	0	0	0.59	0.99	7.46	2.83	2.66
21	1	1	0	1	1	0.77	0.75	0.51	0.25	1.35
22	1	1	0	1	0	0.51	0.5	0.26	−0.01	0.99
23	1	1	0	0	1	1.68	2.28	8.59	3.34	4.59
24	1	1	0	0	0	1.48	2.05	8.31	3.17	4.27
25	1	0	1	1	1	0.67	0.36	−2.24	0.14	0.37
26	1	0	1	1	0	0.42	0.13	−2.53	−0.12	0.04
27	1	0	1	0	1	1.57	1.85	5.4	3.2	3.65
28	1	0	1	0	0	1.37	1.62	5.02	3.04	3.35
29	1	0	0	1	1	1.32	1.74	−3.12	0.56	3.44
30	1	0	0	1	0	1.08	1.48	−3.41	0.29	3.08
31	1	0	0	0	1	2.36	3.5	4.49	3.66	6.94
32	1	0	0	0	0	2.16	3.26	4.12	3.5	6.63

TABLE III-D. Average coding performance [% BD-Rate] for Class D

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
1	**0**	**1**	**1**	**1**	**1**	0	0	0	0	0
2	0	1	1	1	0	−0.12	−0.21	−0.22	−0.2	−0.28
3	0	1	1	0	1	−0.42	0.97	10.54	4.34	4.77
4	0	1	1	0	0	−0.45	0.78	10.4	4.25	4.55
5	0	1	0	1	1	1.76	0.38	0.6	0.06	0.92
6	0	1	0	1	0	1.61	0.15	0.33	−0.15	0.63
7	0	1	0	0	1	1.56	1.48	11.05	4.44	5.8
8	0	1	0	0	0	1.51	1.27	10.91	4.35	5.55
9	0	0	1	1	1	0.41	0.1	−1.94	0.09	0.28
10	0	0	1	1	0	0.29	−0.12	−2.27	−0.1	0
11	0	0	1	0	1	0.04	1.12	8.25	4.45	5.14
12	0	0	1	0	0	0.04	0.94	7.97	4.36	4.9
13	0	0	0	1	1	3.19	0.71	−2.68	0.22	2.27
14	0	0	0	1	0	3.05	0.47	−3	0.02	1.97
15	0	0	0	0	1	3.25	1.99	7.49	4.63	7.32
16	0	0	0	0	0	3.25	1.8	7.22	4.54	7.08
17	1	1	1	1	1	0.87	0.01	−0.33	−0.03	−0.19
18	1	1	1	1	0	0.71	−0.27	−0.54	−0.28	−0.52
19	1	1	1	0	1	0.67	−0.12	8.07	2.95	2.97
20	1	1	1	0	0	0.67	−0.3	7.82	2.82	2.69
21	1	1	0	1	1	2.56	0.34	0.24	−0.01	0.64
22	1	1	0	1	0	2.39	0.06	0.01	−0.27	0.29
23	1	1	0	0	1	2.53	0.31	8.62	2.98	3.84
24	1	1	0	0	0	2.5	0.12	8.37	2.85	3.54
25	1	0	1	1	1	1.28	0.07	−2.38	0.04	0.06
26	1	0	1	1	0	1.13	−0.21	−2.7	−0.21	−0.27
27	1	0	1	0	1	1.1	−0.01	5.8	3.03	3.28
28	1	0	1	0	0	1.13	−0.17	5.43	2.91	3.01
29	1	0	0	1	1	4.02	0.63	−3.11	0.12	1.97
30	1	0	0	1	0	3.86	0.35	−3.43	−0.14	1.62
31	1	0	0	0	1	4.19	0.76	5.05	3.13	5.34
32	1	0	0	0	0	4.21	0.58	4.67	3.01	5.05

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3062938, IEEE Access

TABLE III-E. Average coding performance [% BD-Rate] for Class E

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
1	**0**	**1**	**1**	**1**	**1**	0	0	0	0	0
2	0	1	1	1	0	−0.06	−0.06	−0.08	−0.07	−0.1
3	0	1	1	0	1	2.98	2.26	8.85	3.46	4.65
4	0	1	1	0	0	2.89	2.19	8.72	3.41	4.56
5	0	1	0	1	1	2.14	2.24	1.28	0.42	2.7
6	0	1	0	1	0	2.08	2.17	1.2	0.35	2.59
7	0	1	0	0	1	5.39	4.76	10.31	3.86	7.46
8	0	1	0	0	0	5.3	4.69	10.18	3.81	7.37
9	0	0	1	1	1	0.59	0.74	−2.14	0.28	0.68
10	0	0	1	1	0	0.53	0.67	−2.23	0.21	0.58
11	0	0	1	0	1	3.64	3.1	6.4	3.69	5.31
12	0	0	1	0	0	3.54	3.02	6.29	3.63	5.22
13	0	0	0	1	1	3.35	3.8	−2.61	0.75	4.96
14	0	0	0	1	0	3.29	3.73	−2.69	0.67	4.85
15	0	0	0	0	1	6.77	6.57	5.94	4.15	9.87
16	0	0	0	0	0	6.67	6.49	5.84	4.09	9.77
17	1	1	1	1	1	−0.59	−0.44	−0.25	−0.43	−0.77
18	1	1	1	1	0	−0.64	−0.49	−0.34	−0.5	−0.85
19	1	1	1	0	1	1.5	1.16	7.5	2.25	2.77
20	1	1	1	0	0	1.43	1.11	7.42	2.21	2.71
21	1	1	0	1	1	1.53	1.77	1	−0.03	1.86
22	1	1	0	1	0	1.48	1.72	0.91	−0.11	1.77
23	1	1	0	0	1	3.86	3.6	8.9	2.6	5.47
24	1	1	0	0	0	3.78	3.55	8.82	2.55	5.39
25	1	0	1	1	1	0	0.3	−2.53	−0.18	−0.11
26	1	0	1	1	0	−0.05	0.25	−2.63	−0.25	−0.2
27	1	0	1	0	1	2.16	1.99	4.95	2.46	3.41
28	1	0	1	0	0	2.09	1.95	4.89	2.42	3.34
29	1	0	0	1	1	2.73	3.33	−2.99	0.26	4.1
30	1	0	0	1	0	2.68	3.27	−3.09	0.18	4
31	1	0	0	0	1	5.22	5.38	4.49	2.86	7.82
32	1	0	0	0	0	5.14	5.32	4.42	2.81	7.74

TABLE III-F. Average coding performance [% BD-Rate] for Class F

#	SCL	SAO	DB	RDOQ	TrSk	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
1	**0**	**1**	**1**	**1**	**1**	0	0	0	0	0
2	0	1	1	1	0	3.9	4.68	3.38	5.52	4.29
3	0	1	1	0	1	2.33	1.11	5.39	2.2	2.58
4	0	1	1	0	0	6.65	6.21	10.24	9.23	7.43
5	0	1	0	1	1	1.11	1.3	0.32	−0.48	1.55
6	0	1	0	1	0	5.13	6.11	3.73	5.8	5.9
7	0	1	0	0	1	3.53	2.56	5.74	2.46	4.17
8	0	1	0	0	0	7.91	7.74	10.6	9.5	9.09
9	0	0	1	1	1	1.36	1.56	0.7	0.95	1.72
10	0	0	1	1	0	5.43	6.59	3.15	7.95	5.8
11	0	0	1	0	1	3.7	2.77	6.43	3.93	4.29
12	0	0	1	0	0	8.08	8.06	10.04	11.67	8.86
13	0	0	0	1	1	2.56	3.19	0.27	1.26	4.07
14	0	0	0	1	0	6.77	8.38	2.75	8.29	8.31
15	0	0	0	0	1	5.02	4.61	5.99	4.25	6.76
16	0	0	0	0	0	9.52	10.06	9.64	12.02	11.51
17	1	1	1	1	1	−0.29	−0.11	−0.25	−0.82	−0.2
18	1	1	1	1	0	3.55	4.63	3.2	5.39	4.06
19	1	1	1	0	1	1.06	0.37	4.18	1.35	1.45
20	1	1	1	0	0	5.4	5.54	8.99	8.38	6.26
21	1	1	0	1	1	0.8	1.18	0.05	−0.58	1.31
22	1	1	0	1	0	4.76	6.03	3.55	5.64	5.63
23	1	1	0	0	1	2.25	1.8	4.54	1.59	2.98
24	1	1	0	0	0	6.66	7.05	9.45	8.64	7.86
25	1	0	1	1	1	1.04	1.46	0.42	0.84	1.48
26	1	0	1	1	0	5.07	6.55	2.89	7.77	5.54
27	1	0	1	0	1	2.39	1.96	5.2	3.03	3.09
28	1	0	1	0	0	6.82	7.44	8.81	10.76	7.61
29	1	0	0	1	1	2.24	3.05	0	1.12	3.78
30	1	0	0	1	0	6.41	8.3	2.49	8.08	8.01
31	1	0	0	0	1	3.68	3.77	4.77	3.32	5.5
32	1	0	0	0	0	8.24	9.42	8.42	11.09	10.21

TABLE IV. Average relative CPU encoding time increase/decrease [%], when enabling/disabling a single coding tool from the default encoder configuration (negative values mean time savings)

	QP 22	QP 27	QP 32	QP 37	QP 42	Avg.
SCL on	7.41	8.44	6.57	5.18	3.47	6.21
SAO off	−0.3	−0.21	−0.53	−0.37	−0.58	−0.4
DB off	−0.22	−0.18	−0.41	−0.24	−0.26	−0.26
RDOQ off	−16.56	−10.77	−6.4	−1.89	1.5	−6.82
TrSkp off	−15.78	−14.83	−14.24	−13.29	−13.22	−14.27
SBH off	−3.25	−2.72	−2.17	−1.49	−1.03	−2.13

enabling the SCL coding tool (no matter the status of the rest of the coding tools), we obtain average BD-Rate reductions with all objective quality metrics and video classes (from 0.7% with SSIM to 1.4% with PSNR-HVS); (b) when combining the SCL and RDOQ coding tools, the BD-Rate saving increases an additional 0.9% on average for all objective video quality metrics, so both coding tools complement each other in terms of R/D performance; (c) an exception should be noticed with class D video sequences and the SSIM metric, where enabling the SCL coding tool shows an average BD-Rate increase of 0.9%. (d) The best result was provided by the VMAF metric, scoring an average BD-Rate reduction of 3.58% when both in-loop filters are disabled (SAO and DB) and RDOQ is enabled in class A video sequences.

B. DEBLOCKING FILTER

The deblocking filter minimizes the blocking effect produced by the block partitioning of images during the encoding process. By disabling this filter, the blocking artifacts become visible as the QP value increases.

As can be seen in Tables III-A to III-F, better perceptual performance is provided when the DB coding tool is enabled, independently of the status of the rest of coding tools. This general behavior is observed in all video classes with all the objective quality metrics. The average BD-Rate improvement when enabling DB depends on every objective quality metric and also on the video class. For example, the SSIM, MS-SSIM, and PSNR-HVS metrics always report average BD-Rate savings of 1.2%, 1.5%, and 2.4%, respectively. However, VMAF (in all video classes) and VIF (in classes A and B) report BD-Rate savings when disabling the DB and SAO coding tools (between 0.4% and 1.1% BD-Rate savings). With respect to coding complexity, when DB filter is disabled, an average 0.3% reduction of the encoding time is observed, as shown in Table IV.

C. SAO FILTER

The SAO filter is a technique that attempts to minimize the distortion that is mainly introduced by the quantization step.

We find that in most cases, the perceptual metrics get higher BD-Rate values when disabling the SAO filter. This perceptual worsening is more significant in the synthetic video sequences (class F).

However, the VMAF metric gets better (lower) BD-Rate values for all video classes; it obtains even better results than the results obtained by the default configuration when the SAO filter is disabled, especially if the DB filter is also disabled. The BD-Rate reductions range from 1.8% to 4.4%; the best results occur when both in-loop filters are disabled. When working with class F videos, the VMAF BD-Rate reductions are very low when the SAO filter is disabled; they are always under 1%.

The VIF metric shows a similar behavior to VMAF when working with video classes A and B, achieving up to 0.5% BD-Rate savings when SAO is disabled (0.2% on average).

Finally, as with the deblocking filter, when disabling this filter no significant impact on coding complexity is observed, showing also an average encoding time reduction of 0.5%, as shown in Table IV.

D. RATE-DISTORTION OPTIMIZED QUANTIZATION (RDOQ)

The RDOQ algorithm achieves an estimated optimal quantization value that minimizes the Rate-Distortion cost. In this analysis, we have also disabled the RDOQTS parameter, which deactivates the RDOQ calculation for blocks marked as Transform Skip.

Although the PSNR metric is used by the RDOQ algorithm to measure distortion, looking at the results, we can see that disabling RDOQ parameters implies a deterioration of BD-Rate values for most of the perceptual metrics and video sequence classes. The benefits of enabling RDOQ are more remarkable for the VMAF metric, where on average, 9.8% BD-Rate savings are achieved by enabling the RDOQ coding tool.

When disabling the RDOQ tool, the coding complexity varies depending on the quantization parameter. The highest encoding time reductions are obtained when low quantization values are used (15.78% reduction at QP=22), as shown in Table IV. As the QP value increases, the encoding time savings are progressively reduced.

E. TRANSFORM SKIP

The TransformSkip and TransformSkipFast parameters are enabled by default, since they are able to obtain great BD-Rate savings for artificial or synthetic videos (those belonging to class F).

In Table III-F, corresponding to synthetic or artificial video sequences, we can observe that disabling transform skip parameters causes a significant deterioration of BD-Rate values: all perceptual metrics get BD-Rate increases ranging from 3.6% to 7%.

If we analyze the other video classes, we can see that all of them have slight BD-Rate reductions when disabling the transform skip parameters; these reductions are mostly close to 0% and are never higher than 0.4%.

When disabling the TransformSkip tool, the average encoding time is significantly affected. As stated in Section II-E, disabling this tool reduces the encoding time by almost 15%, as shown in Table IV.

F. SIGN DATA HIDING

The SignHideFlag coding tool, which is activated by default, enables a data compression technique called Sign Data Hiding that provides an average reduction of 0.6% in BD-Rate, regardless of the rest of the settings, as seen in Section II-F.

This coding tool produces a relatively high reduction in rate compared with the very low distortion that it produces, i.e., the perceived quality reported by the perceptual metrics remains almost the same, but the rate is reduced much more.

Table V shows the BD-Rate values obtained by disabling this parameter in the default configuration. As can be seen, by disabling this algorithm, an average BD-Rate increase from 0.71% to 0.95% is achieved.

Since this technique barely distorts the images, we have not included the column corresponding to this parameter in the results tables. Instead, it has been kept in its default state (enabled). The complete tables, including the results of the analysis of the SignDataHiding coding tool, will be available on the GATCOM research group website [38].

TABLE V. Average coding performance [% BD-Rate] by disabling Sign Data Hiding to default configuration

Class	SSIM	MS-SSIM	VMAF	VIF	PSNR HVSM
Class A	0.93	1.18	0.79	1.18	1.12
Class B	0.84	0.95	0.77	1.04	0.94
Class C	0.88	0.83	0.74	0.84	0.9
Class D	1.02	1	0.68	0.89	0.99
Class E	0.52	0.59	0.6	0.68	0.68
Class F	1.05	1.03	0.67	−0.12	1.08
Average	0.87	0.93	0.71	0.75	0.95

V. DISCUSSION

As shown in the previous section, there are coding tools that are perceptually ranked in the same way using all objective quality metrics in every video sequence class, and other coding tools have different behaviors depending on the video classes and/or the objective quality metric used. So, in this section, we will discuss and analyze in detail the results of the encoding tools, taking into account the relationships among them, the metrics and classes, and the reported perceptual behavior.

As stated previously, enabling the Transform Skip parameter works better in the sequences of class F, whereas for the rest of the classes, disabling it slightly increases the perceptual report given by almost all metrics. So, for classes A to E, we can simplify the analysis of the rest of coding tools by disabling Transform Skip and only enabling it when working with videos of class F.

When enabling the Scaling List (SCL) coding tool, we can see that a better perceptual response is reported by almost all metrics, achieving average BD-Rate savings of 0.9%. The behavior of the SCL coding tool is the one expected, since it is well known in the literature that the use of a CSF-based quantizer, implemented by the HEVC through the scaling list coding tool, improves the subjective quality of decoded video [20].

When disabling the RDOQ coding tool, all objective quality metrics show significative BD-Rate increases in all video classes. The average increment of BD-Rate provided in that case, for all metrics and video sequences, is about 4.16%, rising up to 11.5% for VMAF when working with class A. Therefore, we do not recommend deactivating the RDOQ parameter in any case. As there is for every general rule, there is an exception. In this case, the SSIM metric perceives an average BD-Rate reduction of 0.08% in sequences of class D; this goes up to 0.4% when disabling SCL and enabling both in-loop filters, again showing the inability of SSIM to properly score the class D video sequences when compared with the rest of the objective quality metrics.

From the previous analysis, we have shown that in order to provide a better perceptual quality performance for all objective metrics and video sequence classes, we have to (a) enable the SCL and RDOQ coding tools and (b) only enable the TrSk when working with class F video sequences (artificial or synthetic contents).

So, from now on, we will consider that the SCL and RDOQ coding tools are always enabled, and the TrSk is only enabled for class F video sequences. Under this assumption, we will analyze the behavior of the in-loop filters. In Table VI, we show the BD-Rate results of the SAO and DB configurations for each video class, keeping in mind that the rest of the coding tools are enabled/disabled as mentioned above. We have highlighted the maximum average BD-Rate savings of each quality metric and video class.

TABLE VI. BD-Rate evaluation of in-loop filters, with SCL=RDOQ=1 and TrSk=0 (=1 for class F)

SAO	DB	SSIM	MS-SSIM	VMAF	VIF	PSNR-HVSM
Class A						
1	1	**-0,44**	**-0,24**	-0,13	-0,38	**-0,68**
1	0	-0,17	0,77	1,18	-0,66	0,79
0	1	-0,25	-0,04	-2,29	-0,38	-0,43
0	0	-0,06	1,29	**-3,53**	**-0,89**	2,31
Class B						
1	1	**-0,78**	**-0,52**	-0,52	-0,71	**-1,17**
1	0	-0,45	0,85	0,71	-0,68	0,83
0	1	-0,39	-0,05	-2,52	-0,82	-0,74
0	0	-0,31	1,73	**-3,55**	**-1,01**	2,80
Class C						
1	1	**-0,27**	**-0,44**	-0,55	**-0,34**	**-0,56**
1	0	0,51	0,50	0,26	-0,01	0,99
0	1	0,42	0,13	-2,53	-0,12	0,04
0	0	1,08	1,48	**-3,41**	0,29	3,08
Class D						
1	1	**0,71**	**-0,27**	-0,54	**-0,28**	**-0,52**
1	0	2,39	0,06	0,01	-0,27	0,29
0	1	1,13	-0,21	-2,70	-0,21	-0,27
0	0	3,86	0,35	**-3,43**	-0,14	1,62
Class E						
1	1	**-0,64**	**-0,49**	-0,34	**-0,50**	**-0,85**
1	0	1,48	1,72	0,91	-0,11	1,77
0	1	-0,05	0,25	-2,63	-0,25	-0,20
0	0	2,68	3,27	**-3,09**	0,18	4,00
Class F						
1	1	**-0,29**	**-0,11**	**-0,25**	**-0,82**	**-0,20**
1	0	0,80	1,18	0,05	-0,58	1,31
0	1	1,04	1,46	0,42	0,84	1,48
0	0	2,24	3,05	0,00	1,12	3,78

As can be seen, the general behavior of in-loop filters has two opposite positions: (a) the SSIM, MS-SSIM, and PSNR-VHS quality metrics provide the best perceptual results in all video classes when both filters are enabled, showing maximum BD-Rate savings of 0.28%, 0.35%, and 0.66%, respectively, and (b) VMAF (classes A to E) and VIF (classes A and B) say just the opposite, showing maximum BD-Rate savings of 3.40% and 0.95%, respectively, when both in-loop filters are disabled. However, it is worth saying that (a) when working with class F videos, there is a consensus between all metrics in enabling both in-loop filters to maximize the BD-Rate savings, (b) the VIF metric changes its scoring, suggesting that the best configuration for video classes C, D, and E is the one that enables both in-loop filters, joining the group formed by the SSIMM, MS-SSIM, and PSNR-VHS metrics, and (c) with respect to the VMAF metric, we have noticed that it is able to report average BD-Rate savings of 2.53% when only the DB filter is enabled and 0.45% when both filters are enabled.

Although all the objective quality metrics are designed to assess the quality in a way as close as possible to the way that the HVS does, each one uses a different approximation. Some of them perform the subband decomposition inspired by complex HVS models, while others extract structural information from the viewing field or even use the spatio-temporal statistical patterns found in signals captured from the visual field for which the HVS is adapted. Therefore, as we can see in this study, we obtain different quality assessments depending on the metric. In cases where all metrics report BD-Rate variations in the same direction, the conclusion is straightforward, but when the metrics' reports are opposite, a subjective test to validate the results is suggested.

In order to advance a preliminary subjective evaluation that sheds light on the metrics controversy around the in-loop filters' behavior, we have performed a simple subjective test with one class A video sequence. Class A has the highest BD-Rate differences (3.4%) found between the two options: enabling or disabling both in-loop filters (see Table VI). We have chosen a frame of a video sequence where the difference between the VMAF R/D curves of the two options is maximum. We have found that frame 22 of the Traffic_2560x1600_30 video sequence shows a 5.25% BD-Rate reduction when disabling both filters, taking as reference the configuration with the filters enabled. We have observed just noticeable perceptual differences at QPs 37 and 42 with respect to the original frame. These artifacts are more noticeable when the in-loop filters are disabled. In Figure 5, we show a cropped area of frame 22 (encoded with QP = 42), where the blocking artifacts are clearly visible when

(a) Original

(b) With in-loop filters

(c) Without in-loop filters

FIGURE 5. Comparison of cropped section of frame 22 of sequence Traffic (2560x1660) encoded with QP = 42, at $8.5Mbps$

disabling both filters.

We have to mention that the results given by the VMAF are not biased by the image content or by the frame size, as its results are consistent through different frame sizes and content. A good correlation with the DMOS and MOS values of the VMAF has been reported [35]–[37], showing that it can be considered a robust metric. Notice that although the behavior shown in Figure 5 seems to say that the VMAF metric does not correctly assess the perceived quality when both filters are disabled, it shows good results when only the DB filter is enabled. Although this observation does not mean that in-loop filters should always be enabled, a more detailed and carefully designed subjective evaluation test should be performed to determine the best in-loop filter configuration for the A to E video sequence classes.

Finally, another performance metric we may use to assess the most proper coding tool configuration is their contribution to the overall HEVC coding complexity. In Table IV we have shown the time profiling results of each individual coding tool under study, showing their impact on the overall HEVC encoding complexity.

If we enable all coding tools but the TrSk (the best R/D perceptual configuration) the corresponding HEVC overall complexity will be reduced in an 14.27%, on average, when compared to the default HEVC configuration. If we decide to also disable the in-loop filters, SAO and DB, we will get an additional coding time reduction of 0.66%.

VI. CONCLUSIONS

In this article, we have analyzed how the HEVC coding configuration parameters impact the perceptual rate-distortion. To do so, we have used the whole video sequence set defined in the HEVC common test conditions reference and obtained the Bjøntegaard Delta Rate (BD-Rate) measurements for a set of perceptual metrics widely used by the research community. Then, we analyzed how each HEVC coding tool impacts the perceptual BD-Rate and how this relates to other coding tools.

After analyzing the results provided by the set of HEVC coding tools under evaluation, we have arrived at the following conclusions:

a) The coding tools with the highest impact on the overall perceptual quality performance are RDOQ and SCL for all metrics reported in this study, so they should be always enabled.
b) TrSk should be enabled when working with class F videos (artificial, synthetic video contents), as significant perceptual gains are reported. However, for the rest of the video classes, it is slightly better to disable this coding tool. By disabling TrSk, the overall coding time is reduced by 15%.
c) The in-loop filters, SAO and DB, show opposite behaviors when working with video classes A to E, where
   i) one set of metrics (SSIM, MS-SSIM, and PSNR-HVS) implies that both filters should be enabled to maximize the perceptual BD-Rate savings,
   ii) VMAF implies that both filters should always be disabled, and
   iii) VIF shows the same results as VMAF for classes A and B, but for classes C, D, and E, it goes in the same direction as the other metrics.

The recommended HEVC coding tools configuration that will maximize the perceptual R/D should enable both SCL and RDOQ and disable TrSk (enabling it only with class F videos). As discussed in the previous section, there is no agreement with respect to the in-loop filters (SAO and DB). Three alternatives exist: (a) enable both filters, (b) disable them, and (c) only enable DB filter. The encoding complexity of both filters is low; therefore, their complexity does not help so much to take a firm decision. So, to determine the best option, we need to design specific subjective tests, taking into account the target video classes, to decide which one should be used.

The data presented in this article is intended to help other researchers to determine the best encoder configuration, depending on the type of sequence to be coded, to maximize the perceptual rate-distortion performance, taking into account the coding tools complexity. Also, it can be useful to choose the most appropriate perceptual metric to be used in the design of subjective tests.

As future work, we plan to extend this study by including more HEVC coding tools and perform exhaustive subjective tests to determine the perceptual-based settings that should be configured in the HEVC encoder to maximize the R/D quality performance.

REFERENCES

[1] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding (Signal Processing and Communications)*. USA: CRC Press, Inc., 2005.

[2] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High dynamic range video compression by intensity dependent spatial quantization in HEVC," *2013 Picture Coding Symposium (PCS)*, pp. 353–356, 2013.

[3] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High Dynamic Range Video Compression Exploiting Luminance Masking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 950–964, 2016.

[4] M. Naccari, M. Mrak, D. Flynn, and A. Gabriellini, "Improving HEVC compression efficiency by intensity dependant spatial quantisation," in *10th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Document: JCTVC-J0076, Jul 2012.

[5] P. Ren, J. Wang, J. Snyder, X. Tong, and B. Guo, "Pocket Reflectometry," *ACM Trans. Graph.*, vol. 30, no. 4, Jul 2011. [Online]. Available: https://doi.org/10.1145/2010324.1964940

[6] B. Girod, *What's Wrong with Mean-Squared Error?* Cambridge, MA, USA: MIT Press, 1993, p. 207–220.

[7] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, Nov 2003, pp. 1398–1402 Vol.2.

[9] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr 2004.

[10] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommunication Systems*, vol. 49, pp. 35–48, 2012.

[11] F. Bossen, "Common test conditions and software reference," in *11th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Doc. JCTVC-K1100, Oct 2012.

[12] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, Jun 2008.

[13] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara. (2016, Jun) Toward a Practical Perceptual Video Quality Metric. Netflix TechBlog. [Online]. Available: https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652

[14] F. Bossen, D. Flynn, K. Sharman, and K. Sühring, "HM Software Manual," Joint Collaborative Team on Video Coding (JCTVC) of ITUT SG16 WP3 and ISO-IEC JTC1-SC29-WG11, Tech. Rep.

[15] M. O. Martínez-Rach, "Perceptual image coding for wavelet based encoders," Ph.D. dissertation, Universidad Miguel Hernández de Elche, Dec 2014. [Online]. Available: https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1128660#

[16] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, Jul 1974.

[17] N. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," *IEEE Trans. on Communications*, vol. 33, no. 6, pp. 551–557, Jun 1985.

[18] B. Chitprasert and K. R. Rao, "Human visual weighted progressive image transmission," *IEEE Trans. on Communications*, vol. 38, no. 7, pp. 1040–1044, Jul 1990.

[19] M. Wien, *High Efficiency Video Coding: Coding Tools and Specifications*, ser. Signals and Communication Technology. Springer, 2015.

[20] M. Haque, A. Tabatabai, and Y. Morigami, "HVS model based default quantization matrices," in *7th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Document: JCTVC-G880, Nov 2011.

[21] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 614–619, Jul 2003.

[22] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. Van der Auwera, "HEVC Deblocking Filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1746–1754, Dec 2012.

[23] C. Fu, E. Alshina, A. Alshin, Y. Huang, C. Chen, C. Tsai, C. Hsu, S. Lei, J. Park, and W. Han, "Sample Adaptive Offset in the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1755–1764, Dec 2012.

[24] G. Bjontegaard, "Calculation of average PSNR differences between RD-Curves," in *Proc. of the ITU-T Video Coding Experts Group - Thirteenth Meeting*, Apr 2001.

[25] J. Stankowski, C. Korzeniewski, M. Domański, and T. Grajek, "Rate-distortion optimized quantization in HEVC: Performance limitations," in *2015 Picture Coding Symposium (PCS)*, 2015, pp. 85–89.

[26] L. Prangnell, "Visually lossless coding in HEVC: A high bit depth and 4:4:4 capable JND-based perceptual quantisation technique for HEVC," *Signal Processing: Image Communication*, vol. 63, pp. 125–140, 2018.

[27] J. Kao, M. A. Hashemi, X. Xiu, Y. Ye, Y. He, and J. Dong, "Improved transform skip mode for HEVC screen content coding," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Nov 2015, pp. 504–509.

[28] G. Clare, F. Henri, and J. Jung, "Sign Data Hiding," in *7th Meeting of the Joint Collaborative Team on Video Coding (JCT-VC)*. Document: JCTVC-G271, Nov 2011.

[29] F. N. Fritsch and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," *SIAM Journal on Numerical Analysis*, vol. 17, no. 2, pp. 238–246, 1980. [Online]. Available: https://doi.org/10.1137/0717021

[30] Fraunhofer Institute for Telecommunications. (2018, Sep) HM Reference Software Version 16.20. [Online]. Available: https://hevc.hhi.fraunhofer.de/trac/hevc/browser/tags/HM-16.20

[31] H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessmen," in *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005, pp. 23–25.

[32] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in *Proceedings of the third international workshop on video processing and quality metrics*, vol. 4, 2007.

[33] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 935–949, 2011.

[34] Recommendation ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," *International Telecommunication Union*, Apr 2008.

[35] R. Rassool, "VMAF reproducibility: Validating a perceptual practical video quality metric," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun 2017, pp. 1–2.

[36] J. Gutiérrez, L. Krasula, P. Le Callet, Z. Li, and I. Katsavounidis, "VMAF Fframework performance on UHD videos," in *VQEG Meeting 2017*, Los Gatos, CA, USA, May 2017.

[37] C. Lee, S. Woo, S. Baek, J. Han, J. Chae, and J. Rim, "Comparison of objective quality models for adaptive bitstreaming services," in *2017 8th International Conference on Information, Intelligence, Systems Applications (IISA)*, 2017, pp. 1–4.

[38] Universidad Miguel Hernández de Elche, "Grupo de Arquitectura y Tecnología de Computadores," http://atc.umh.es/gatcom/Ficheros/overall_results.zip. [Online]. Available: http://atc.umh.es/gatcom/

JAVIER RUIZ ATENCIA received his B.S. degree in Telecommunications Technology Engineering from Miguel Hernández University (Spain) in 2014. Thanks to the SICUE mobility programme, he completed his last year at Polytechnic University of Catalonia (UPC), where he received distinction on his final project. In 2019, he received his M.S. in Telecommunications Engineering from Miguel Hernández University. In 2012, he participated in the creation of the IEEE UMH Student Branch. Between 2014 and 2019, he worked in the R&D department of an important energy efficiency company. He is currently a PhD student at the Computer Engineering Department of Miguel Hernández University. His research includes perceptual compression techniques applied to the latest image and video standards.

OTONIEL LÓPEZ GRANADO received his M.S. in Computer Science from the University of Alicante (Spain) in 1996. Between 1997 and 2003, he worked as a programmer analyst in an important industrial informatics firm. In 2003, he joined the Computer Engineering Department at Miguel Hernández University (UMH), Spain, as an assistant professor. Then, he received his Ph.D. degree in Computer Science in 2010. In 2012, he was promoted to associate professor. Currently, he leads the GATCOM research group (atc.umh.es) at Miguel Hernández University. His research and teaching activities are related to multimedia networking (audio/video coding and network delivery).

MANUEL P. MALUMBRES received his B.S. in Computer Science from the University of Oviedo (Spain) in 1986 and the M.S. and Ph.D. degrees in Computer Science from the Technical University of Valencia (UPV) in 1991 and 1996, respectively. He is the author of more than 200 conference and journal publications and several networking books for undergraduate CS courses. Currently, his research and teaching activities are related to multimedia networking (image/video coding and network delivery), wireless network technologies (MANETs, VANETs, and WSNs), and acceleration schemas for multimedia applications (multi/many-threads, GPUs, FPGAs).

MIGUEL ONOFRE MARTÍNEZ-RACH received his M.S. degree in Computer Science from the University of Alicante in 1996. He worked for a multinational French-owned computer company as a data warehouse analyst. In 2003, he entered Miguel Hernández University, where at present he works as an associate professor in the Computer Engineering Department. He obtained his Ph.D. degree (extraordinary doctorate award) from Miguel Hernández University in 2014. His teaching subjects include Operating Systems and Mobile Programing. His research subjects are related to image and video compression, specifically with perceptual coding. Currently his research is focused on the perceptual enhancement of the HEVC standard.

GLENN VAN WALLENDAEL received the M.Sc. degree in applied engineering from the University College of Antwerp, Belgium, in 2006, and the M.Sc. degree in engineering from Ghent University, Belgium, in 2008. He is currently pursuing the Ph.D. degree with IDLab, Ghent University—imec, with the financial support of the Research Foundation—Flanders (FWO), where he is currently working in the same group as a Post-Doctoral Researcher. His main topics of interest are video compression, including scalable video compression and transcoding.

• • •