

# Quality comparison of the HEVC and VP9 encoders performance

Pedro Fernandes<sup>1</sup> · Marco V. Bernardo<sup>1</sup> ·  
António M. G. Pinheiro<sup>1</sup> · Paulo T. Fiadeiro<sup>2</sup> ·  
Manuela Pereira<sup>1</sup>

Received: 18 December 2015 / Revised: 20 June 2016 / Accepted: 27 June 2016  
© Springer Science+Business Media New York 2016

**Abstract** This paper reports a comparison between two recent video codecs, namely the HEVC and the VP9, using High Definition Video Sequences encoded with different bit rates. A subjective test for the evaluation of the provided Quality of Experience is reported. The video sequences were shown to a panel of subjects on a High Definition LED display and the subjective tests were performed using a Single Stimulus Methodology. The results shown that the HEVC encoder provides a better visual quality on low bit rates than the VP9. Similar performance was obtained for visually lossless conditions, although the HEVC requires lower bit rates to reach that level. Moreover, the correlation of the subjective evaluation and three tested objective metrics (PSNR, SSIM, and FSIM) revealed a good representation of the subjective results, particularly the SSIM and the FSIM metrics.

**Keywords** Video codecs · Visual communication · Image quality · Quality assessment

---

✉ Marco V. Bernardo  
mbernardo@ubi.pt

Pedro Fernandes  
pedro.manuel.fernandes@ubi.pt

António M. G. Pinheiro  
pinheiro@ubi.pt

Paulo T. Fiadeiro  
fiadeiro@ubi.pt

Manuela Pereira  
mpereira@ubi.pt

<sup>1</sup> Instituto de Telecomunicações (IT), Universidade da Beira Interior, Rua Marquês d' Ávila e Bolama, 6201-001, Covilhã, Portugal

<sup>2</sup> Fiber Materials and Environmental Technologies (FibEnTech), Universidade da Beira Interior, Rua Marquês d' Ávila e Bolama, 6201-001, Covilhã, Portugal

# 1 Introduction

In recent years the number of multimedia information consumers did not stop to grow, requiring new video encoding technology and demanding improved services. When considering visual applications, the common term Quality of Service (QoS) is not enough to provide a proper description of the performance of a system. In contrast, the Quality of Experience (QoE) evaluation involves the subjective factors of the end user. For this reason, the QoE is more appropriate for the performance evaluation of real systems. The authors in [3] identified general descriptive characteristics of experienced quality for multimedia systems, and was concluded that the combination of quality components requires the execution of subjective quality ratings of potential users. The subjective evaluation of audio and visual quality is considered the most accurate method that considers the human perception [13]. Particularly, the visual 2D subjective quality evaluation, according to standardized methods, has been widely studied in recent years. Therefore, the International Telecommunication Union (ITU) has issued several recommendations on the methodology for subjective quality evaluation, including the ITU-R BT.500 [12] and the ITU-R BT.2022 [13]. The last one reports similar methods to [12], but take into account the subjective quality evaluation of emerging multimedia data, such as Higher Definition TV formats, High Dynamic Range (HDR), or 3DTV.

One of the main factors that influence the video quality is the encoding method and therefore the QoE perceived by the consumers. Particularly, on video streaming, the encoding method has a strong influence on the required transmission speed, or in the amount of required storage. Therefore, the current network distribution or storing models require a balance between the bit rate and quality. Most of the used video encoding methods were established by the Moving Picture Experts Group (MPEG) [22] that defined the High Efficiency Video Coding (HEVC) standard [26]. However, Google decided to stop using the MPEG ISO standards in their video applications and has created a new free access codec. The latest version of this encoder family is the VP9 [29] that is a successor of the VP8 [30]. The VP9 was created to improve the compression rate of the VP8, especially for HD content [1], and to compete with the existing ISO standards.

Currently, little literature can be found considering subjective or objective quality evaluation of the HEVC and the VP9, since both encoders are still recent. The authors in [21] present an objective evaluation of the HEVC encoding and compare it with its predecessors. In [5, 31] and [7] the authors present results on a subjective quality of the HEVC encoder. In these works the HEVC is compared with its predecessor H.264/MPEG-4 Advanced Video Coding (AVC). Comparisons between the HEVC and the VP9 encoders using objective evaluation was presented in [6], using the Peak Signal-to-Noise Ratio (PSNR). Moreover, a subjective evaluation was presented on the recent works [23] using a Ultra High Definition (UHD) television broadcast scenario. The authors computed the Mean Opinion Scores (MOS) using a Double Stimulus Impairment Scale (DSIS) method to compare the quality of the different encoding schemes. These works will be detail in the related work section.

This paper reports a subjective and objective quality evaluation, comparing the HEVC and the VP9 encoders using a set of HD videos, that allows to verify what is the standard that provides the best relation between bit rate and subjective quality trade off. For the subjective quality evaluation the Absolute Category Rating with Hidden Reference (ACR-HR) standard test methodology [16] was chosen. This methodology was selected because it approximates to the real viewing conditions of multimedia systems consumers, that do not

have any reference. Moreover, the objective quality evaluation was compared with existing results in the literature. Furthermore, the correlation between the results of the subjective quality evaluation and the results of three objective metrics is also presented.

## 2 Related work

The number of publications analyzing the performance of the HEVC and VP9 is still limited, since both encoders are still recent.

The coding efficiency that can be achieved by using the HEVC codec, and its major predecessors was analyzed by Ohm et al. [21]. This analysis includes the H.262/MPEG-2 Video [8, 10, 19], the H.263 [14], the MPEG-4 Visual [11], and the H.264/MPEG-4 AVC [15, 27, 32]. The profiles that provide the best coding efficiency were chosen. All videos were encoded using the same Lagrangian-based optimization technique that was applied for mode decision, motion estimation, and quantization. The authors used a Bjntegaard-Delta Bit Rate (BD-BR) measurement method as evaluation criterion. The results indicate that the emerging HEVC standard clearly outperforms its predecessors in terms of coding efficiency. The bit rate savings for the low bit rate range are generally somewhat higher than the average savings. The DSIS was used for the subjective quality evaluation. Results of the subjective tests were provided for the HEVC and the AVC comparison. These results indicated that about 50 % of bit rate reduction could be achieved for the test set of video sequences. The subjective benefit for the HEVC seems to exceed the benefit measured using the PSNR.

Hanhart et al. [7] provide subjective quality evaluation comparing the performance of the HEVC (HM 6.1.1) and the H.264/MPEG-4 AVC (JM 18.3) on 4K/QFHD video content. The Random Access configuration was selected. The configuration parameters for the AVC and the HEVC were selected such that similarity was ensured between the two codecs. The evaluation was performed on natural and synthetic contents with different spatio-temporal characteristics. The DSIS method, Variant II, with a continuous impairment scale was chosen to perform the subjective quality evaluation experiments. The results show that, especially for lower bit rates, the performance of the HEVC exhibits a substantial quality improvement compared to the AVC. In most cases, a significant difference is observed between the HEVC and the AVC for a similar bit rate. It was also noticed that the BD-PSNR under estimates the actual bit rate reduction.

Garcia et al. [5] present a subjective comparison between the H.264/MPEG-4 AVC (JM18.3) and the HEVC (HM 6.0) encoding standards in accordance with the DSIS for mobile video services. Test environments are based on the higher resolution ( $640 \times 360$  pixels) of smart phone LCD and expected cellular bit rates, such as 200 kbps, considered a constrained transmission bandwidth, and 400 kbps, considered a good transmission rate for mobile resolution encoded sequences. The AVC encoder was configured to closely emulate the HEVC coding based on the HM-like configurations available in JM 18.3, to reduce configuration variability between the two video encoding protocols. Subjective feedback shows that both the encoding methods are adequate at 400 kbps constant bit rate. A consumer experience gap was observed for 200 kbps constant bit rate. Significantly less, the AVC subjective quality is noticed with video sequences that have multiple objects moving and no single point of visual attraction. The video sequences with single points of visual attraction or a few moving objects tended to have the AVC encoded video on par with the HEVC encoded video.

Weerakkody et al. [31] present results and analysis of the HEVC standard (HM-12.1) compared with its predecessor the H.264/MPEG-4 AVC (JM-18.5), with the focus on UHD video content. A comparison and analysis of the PSNR and the subjective quality evaluation results were presented. The sequences were compressed using Random Access profile. The HEVC used the Main profile while AVC used the High profile. Both software encoders were configured to use similar referencing structures and similar rate-distortion optimization. The Degradation Category Rating (DCR) method was used for this test under the scheme quality evaluation. Results shown that the HEVC could provide the same visual quality as AVC for UHD content at well below half the bit rate of the latter, surpassing the performance expected at the launch of the HEVC standard development process. In all tested points, the bit rate savings for equal MOS was higher than the bit rate savings for equal PSNR.

Grois et al. [6] presents a performance comparison of the H.264/MPEG-4 AVC (x264 encoder) and the HEVC (HM 10.0) as well as the video coding scheme VP9. Results were obtained for a whole test set of video sequences by using similar encoding configurations for all three examined representative encoders. The HEVC encoder provides significant gains in terms of coding efficiency compared to both VP9 and x264 encoders. It is also observed that the bit rate savings, on average, are increasing along with an increase of quantization parameters for both VP9 and x264 encoders.

Rerabek et al. [23] present the performance analysis and mutual comparison of video coding standard HEVC (HM 15.0) with the video coding scheme VP9 (version 1.2.0), using a UHD television broadcast scenario. Authors included the H.264/MPEG-4 AVC (JM 18.6) into the evaluation to serve as a comparison baseline. The Random Access configuration was selected for both JM and HM reference encoders. Four different UHD contents with various spatio-temporal characteristics have been encoded at five bit rates for each codec, and the subjective quality scores related to content, codec and quality have been collected. The subjective evaluation uses the DSIS method, Variant II, with a continuous impairment scale. The results indicate a dominance of the HEVC based encoding algorithm in comparison to other alternatives. The subjective scores shown a higher average bit rate reduction than the one obtained with the PSNR. In addition, the VP9 shown competitive results for synthetic content and bit rates that correspond to operating points for transparent or close to transparent quality video.

### 3 Subjective quality evaluation

#### 3.1 Laboratory

The laboratory used for subjective quality assessment is in the Optics Center at Universidade da Beira Interior (UBI), which is compliant with the ITU recommendation [12]. The experiment was prepared following the ITU recommendation [13]. In every session, groups of three subjects simultaneously visualized the displayed videos. They were seated in one row perpendicular to the monitor at a distance equal to 187 cm, which is 3.2 times the image height (58.53 cm). The videos used in this study were displayed on a 47 LG TV color display, model 47LA860V, and were visualized inside a dark room with dim illumination.

#### 3.2 Selected codecs

In this work, the HEVC [26] and the VP9 [29] codecs were considered. The HEVC is the standard video encoder that succeeds the MPEG-4/H.264, that appeared in 2003. The

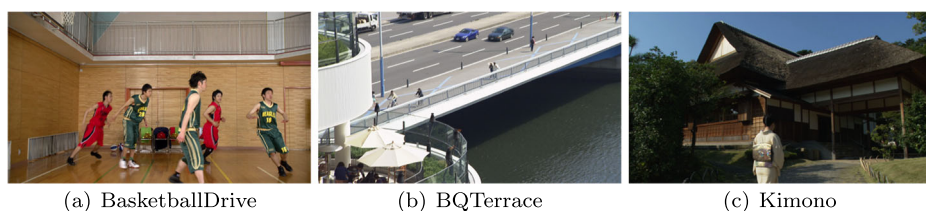
MPEG-4/H.264 was regarded as the best codec for video compression since its introduction on the market. In order to further improve the performance of H.264, mostly for higher resolutions videos, the HEVC standard was developed. This encoder presents higher compression rates, mostly for HD and UHD video. According to its creators this video encoder aims to increase video quality and double the rate compared to H.264 compression. It already supports high resolutions up to  $8192 \times 4320$  pixels which is approximately 4 times larger than the current HDTV with  $1920 \times 1080$  pixels. In the test case for intra-only coding, each picture in a video sequence shall be encoded as an Instantaneous Decoder Refresh (IDR) picture, and no temporal reference pictures shall be used. For the low-delay coding only the first picture in a video sequence shall be encoded as an IDR picture, and the other successive pictures shall be encoded as Generalized P or B-pictures (GPB). For the random-access coding, a hierarchical B structure shall be used for encoding, since it is the one with best rate/quality performance. The first intra picture of a video sequence shall be encoded as an IDR picture and the other intra pictures shall be encoded as non-IDR intra pictures. The pictures located between successive intra pictures in display order shall be encoded as B-pictures.

Considered to be the main opponent of the HEVC encoder, the VP9 is an open source video encoder developed by Google, with the main purpose of reducing the final video bit rate. Hence, lower bandwidth consumption will result, a faster loading of the video will be provided, and also less disk space will be required for storing. The VP9 defines the profile 0 and profile 1. The first supports 4:2:0, while the profile 1, which is optimized for hardware, supports 4:4:4. The VP9 only uses progressive encoding, and divides each picture in a video sequence into blocks of size  $64 \times 64$  pixels, which are called superblocks. These superblocks are processed from left to right and top to bottom, as most video encoders do, and can be also subdivided into sub-blocks of size  $4 \times 4$  pixels. This subdivision is done through recursive subdivision, as in the HEVC encoder. However, unlike the HEVC, a subdivision can be done horizontally or vertically.

### 3.3 Video sequences dataset

The dataset of the testing video sequences was created from three original video sequences after encoded with the HEVC and VP9. These original videos sequences, with a representative frame presented in Fig. 1, were obtained from a repository provided by JCT-VC [20]. In Table 1 are described the identification, resolution and frame rates of these original videos sequences.

In terms of video characteristics, it is important to note that the BasketballDrive sequence presents a fast motion (camera and objects) but the scenario is constrained. Considering the BQTerrace sequence, it presents a medium motion (objects and camera) but the scenario is wide and varies. Finally, the Kimono sequence is a slower motion sequence, which does not



**Fig. 1** Representative frames of each original video sequences

**Table 1** Description of each original video sequences

Video Sequence	Resolution	Frame rate
BasketballDrive	1920 × 1080	50 fps
BQTerrace	1920 × 1080	60 fps
Kimono	1920 × 1080	24 fps

present many details and is moderately constrained. All of these selected video sequences use the YUV 4:2:0 format.

### 3.3.1 General codecs configuration

To generate the encoded videos one configuration is applied to each of the used codecs.

For the HEVC codec, it was used the most popular software encoder, the HEVC Test Model (HM),<sup>1</sup> with the 13.0<sup>2</sup> version. The profile random-access was used for video encoding since it is the one with best rate/quality performance. The values used for parameterization were the ones proposed on the configuration file by the Joint Collaborative Team on Video Coding (JCT-VC) [20], and the details can be found in Table 4 presented in Appendix.

For the VP9 codec, it was used the version v1.3.0.<sup>3</sup> The parameters used for this codec were the recommended by [6]. Authors in [6] tested three different configurations, the two-pass best quality settings recommended by the leading VP9 senior developers [2] (configuration 1 and configuration 2), and the most recommended VP8 two pass best quality settings [30] (configuration 3). However, they conclude that all three configurations lead substantially the same performance results. Table 5 in Appendix present this study selected settings for the VP9 encoder that correspond to configuration 1 in [6].

The Quantization Parameter (QP) and the Intra Period Parameter (IPP) were changed in the different tests. Particularly, for the VP9 codec, the parameters *max-q* and *min-q* were used for the QP, while the *kf-min-dist* and *kf-max-dist* were used for the IPP. The parameter *w* defines the width of the video, while the *h* defines the height. More details about the restart parameters can be found on the encoder parameter guide [30].

### 3.3.2 Quantization and intra period configuration

The sequences were encoded and decoded with the general parameterizations presented above, and with QP values of 22, 27, 32 and 37, and IPP values of 24, 32, 56 and 64. All encoded/decoded sequences were analyzed. From this analysis was verified that the values of QP equal 22 and the IPP equal 64 would not be necessary. A QP equal to 22 results in sequences indistinguishable from the originals. The sequences with IPP of 64 and 56 were very similar, thus only the value 56 was used in the experiment. It was also observed that the VP9 result in higher bit rates for these QP than the ones obtained with the HEVC. Thus, a QP equal to 46 was also used for this codec.

<sup>1</sup>[https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware).

<sup>2</sup>[https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/tags/HM-13.0](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-13.0).

<sup>3</sup><http://www.webmproject.org/vp9>.

Hence, in the experiment, QP values of 27, 32, 37 were chosen for the HEVC codec, and the QP values of 32, 37, 46 were chosen for the VP9. The IPP values of 24, 32, and 56 were used for both codecs. Therefore, a total of 54 videos sequences were obtained, plus a sample of each original video, resulting in a dataset with 57 video sequences.

### 3.4 Panel of subjects

The panel of subjects was composed by 30 volunteers, 22 male and 8 female, within the range of 18-34 years old and an average age of  $22.5 \pm 3.2$  years old. The subjects were organized in groups of three for visualization proposes.

All of them were naïve as to the aim of the experiment and all had normal color vision. Informed consent was obtained from all participants and the research was conducted according to the Declaration of Helsinki [33].

### 3.5 Test methodology

The Absolute Category Rating with Hidden Reference (ACR-HR) standard test methodology [16] was chosen for the experiment. In this methodology a sequence of videos is presented to the subjects, one at each time. The sequences are evaluated independently with a classification scale. This methodology was chosen because it approximates real viewing conditions as consumers of a multimedia system do not have any reference.

Prior to the beginning of the test, an experimental protocol was shown to the subjects. Then, two parts composed the test. The first one, called “training period”, allowed the subjects to familiarize with the evaluation procedure. This was followed by the second part, the “test session” itself, which comprises the visualization of the 57 generated videos sequences as described. The “training period” used five random videos selected from the “test session”. Only the answers given by the subjects during the “test session” were considered as results.

To reduce contextual effects, the video sequences visualization order were randomized applying different permutations for each group of subjects. Moreover, the same video content was not shown consecutively. The tests comprise the subjects visualization of a video during approximately 10 seconds followed by a period of 10 seconds with a mid gray image. This last period is used for the video quality evaluation, where the subjects gave a grade between 1 (low quality) and 10 (high quality), as defined in [16].

The test had a global duration of approximately 20 minutes. Longer test sessions are not advisably because the effects of the subjects tiredness would influence the final results. The results of these subjective tests were used to calculate the MOS.

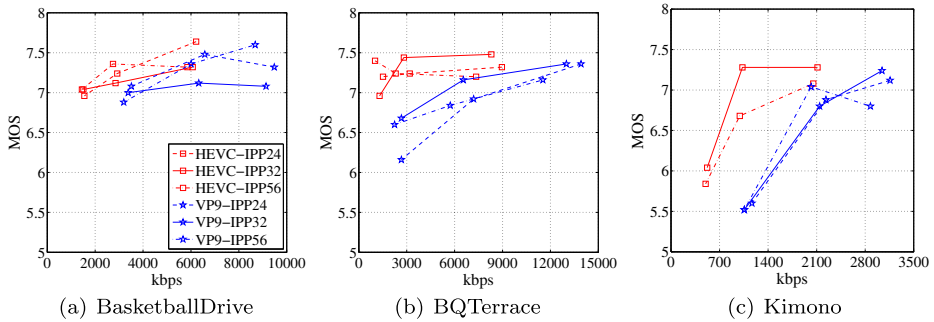
## 4 Results analysis

The analysis of the subjective and objective quality evaluation, is reported in this section for the above test methodology.

### 4.1 Subjective quality evaluation

The subjective results were analyzed for standard outlier detection [4] in order to remove subjects whose scores reveal a strong deviation from the other scores in the same session. Five subjects were considered as outliers, and thus all their scores were removed from the results.





**Fig. 2** MOS of the video sequences as a function of the bit rate (kbps)

The MOS was computed using the subjective evaluation results, providing a numerical indication of the perceived quality from the users perspective. The calculation of the MOS of a video sequence is given by (1).

$$\text{MOS}(j) = \frac{\sum_{i=1}^N s_{ij}}{N}, \quad (1)$$

where  $N$  is the number of subjects, and  $s_{ij}$  is the evaluation of the subject  $i$  for the video sequence  $j$ .

The MOS values obtained for each video sequence used in the subjective tests are presented in Fig. 2 and shows that higher bit rates lead to better ratings for both encoders. Similar performance was obtained for visually lossless conditions. These are similar to the ones obtained with the original sequences (see Table 2). The MOS was always above 5 for these experiments. The lowest grade given by subjects was 5.4 and the highest score was 7.6. The MOS results of the sequences coded at higher bit rate were similar to the MOS results obtained for the original sequences. In fact, these cases are visually near lossless conditions. The MOS values for sequences coded at lower bit rates were always lower, except the BQTerrace sequence coded with HEVC with IPP set to 56. After analyzing this particular case, we verified that in this sequence some textures from peripheral view present less detail. This situation might create a sensation of larger visual comfort to viewers, leading to this unexpected result.

From the analysis of all subjective results can be concluded that for the used parameter values, the HEVC encoder performs better than the VP9 encoder. However, the performance gain is reduced. Low bit rates are more noticeable. For instance, the BQTerrace sequence encoded with a bit rate of approximately 2300 kbps, results in a MOS value of 6.4 with the

**Table 2** Original video sequences (Bit rate, MOS and standard deviation)

Video Sequence	Bit rate (kbps)	MOS	Standard Deviation
BasketballDrive	1215000	7.40	1.65
BQTerrace	1458000	7.44	1.06
Kimono	583200	7.48	1.17



VP9 encoder while the HEVC leads to a MOS of 7.3. The same sequence encoded with a bit rate of almost 10000 kbps, results in a similar evaluation of approximately 7.4 for both encoders. The 7.44 was also the evaluation value obtained for the original BQTerrace sequence, as can be verified in Table 2.

Finally, the two codecs subjective results were compared statistically with two normality tests performed on the sample variables. The Shapiro-Wilk [25] and the Kolmogorov-Smirnov [18] tests both results on a non-normally distributed data. Thus, the non-parametric Wilcoxon test [9] (95 % confidence interval) was performed, and it was concluded that between this two samples there is no statistically significant difference with a  $p$ -value equal to 0.1469 for the each IPP value. However, the resulting bit rates for each IPP value are different. In this case, the bit rate is a non-normally distributed data, and the Wilcoxon test results in a  $p$ -value equal to 0.0039, which for a 95 % confidence interval means that the distributions are statistically different. Hence, the differences in the bit rate *vs* MOS represented in Fig. 2 are statistically different and can be concluded that the HEVC requires lower bit rates than the VP9 for the same subjective quality.

## 4.2 Objective quality evaluation

For objective evaluation the PSNR, Structural Similarity Index (SSIM) [28] and Feature SIMilarity Index (FSIM) [34] metrics were used.

The luma (Y) and chroma (U,V) components were combined for the PSNR metric computation using the (2),

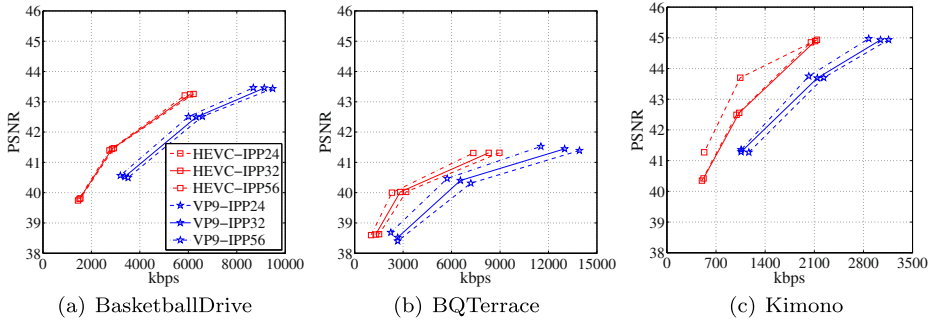
$$\text{PSNR} = \frac{6 \times \text{PSNR}_Y + \text{PSNR}_U + \text{PSNR}_V}{8} \quad (2)$$

The PSNR results are presented in Fig. 3 and are in conformance with the reported in the literature [6, 7, 21]. For the same bit rate the videos coded with the HEVC present a higher PSNR than the same videos coded with the VP9. In general the HEVC coder is 1.5 dB better than the VP9. As can be also seen, the PSNR almost does not change with the tested IPP values variation.

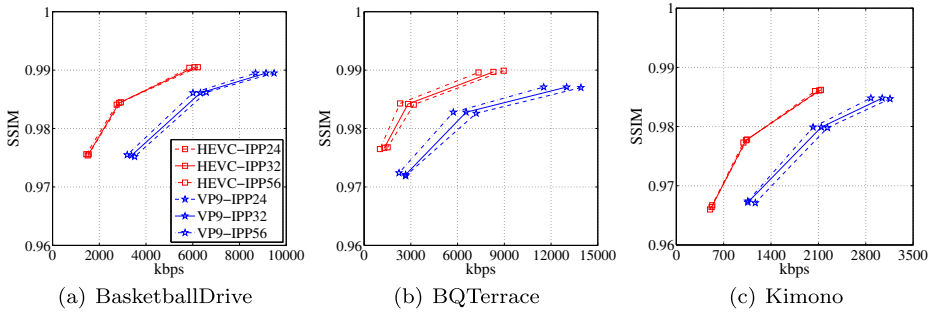
The SSIM [28] and the FSIM [34] are based on the human visual system (HVS) to provide an improved representation of the perceptual quality. The SSIM metric uses structural information from natural images, and evaluates perceptual quality using three spatially local evaluations: mean, variance, and cross-correlation [24]. The FSIM metric is based on the SSIM and adds the comparison of low-level feature sets between the reference and the distorted images [34]. The results obtained with these two metrics are presented in Figs. 4 and 5. The HEVC result in better performance for all the three metrics. Furthermore, the FSIM metric leads to a larger difference between the two encoders, than the other two metrics, particularly in the case of BQTerrace sequence. The Kimono sequence presents the higher PSNR while BQTerrace sequence present the lowest PSNR. These grading is not confirmed by SSIM and FSIM metrics. The BasketballDrive sequence is the one with highest SSIM and FSIM values while the Kimono sequence is the one with lowest values for SSIM and FSIM metrics. The SSIM and FSIM results for the BasketballDrive sequence are very similar, but there are important differences in the other two video sequences.

### 4.2.1 Evaluation of the objective models

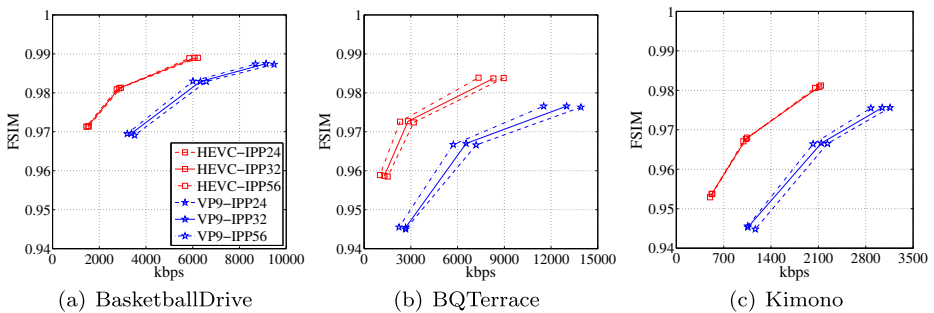
The original subjective results were normalized into  $\text{MOS}_n$  in the range [0, 1], which consists of a simple continuous relation between the original MOS and the coded video



**Fig. 3** PSNR of the video sequences as a function of the bit rate (kbps)



**Fig. 4** SSIM of the video sequences as a function of the bit rate (kbps)



**Fig. 5** FSIM of the video sequences as a function of the bit rate (kbps)

Table 3 Performance of the quality metrics

Codec	Metric	Measures			
		Pearson	Spearman	Outlier Ratio	RMSE
HEVC	PSNR	0.382	0.540	0.000	0.216
	SSIM	0.943	0.827	0.000	0.075
	FSIM	0.902	0.779	0.000	0.097
VP9	PSNR	0.453	0.530	0.033	0.236
	SSIM	0.943	0.876	0.033	0.088
	FSIM	0.886	0.883	0.033	0.124

sequences. In this range, zero represents the original and one represents the lowest quality. The normalization is presented in (3), where  $j$  is the  $j^{th}$  tested video sequence.

$$MOS_n(j) = \frac{MOS(j) - MOS_{min}}{MOS_{max} - MOS_{min}} \tag{3}$$

Then, the non-linear regression suggested in [17], was fitted to the objective metrics results and the mapped  $MOS_n$  values, and restricted to be monotonic over its range. The data  $[MOS_p, MOS_n]$  was fitted using (4) for the regression,

$$MOS_p(j) = b1 + \frac{b2}{1 + e^{(-b3 \times (MR(j) - b4))}} \tag{4}$$

where  $MOS_p$  represents the MOS prediction,  $j$  is the  $j^{th}$  tested video sequence,  $MR$  is the metric result, and  $b1, b2, b3$  and  $b4$  denote the regression parameters, initialized with  $MOS_{min}, MOS_{max}, MR_{min}$  and  $MR_{max}$  respectively.

The metrics performance were evaluated using the common four measures: 1) The Pearson linear correlation coefficients between  $MOS_n$  and  $MOS_p$ , that measures the prediction accuracy of the model; 2) The Spearman rank order correlation coefficient between  $MOS_n$  and  $MOS_p$ , that relates the prediction monotonicity of the model; 3) The Outlier Ratio as a measure of the model consistency prediction; 4) The root mean square error (RMSE). The results of these measures are presented in Table 3.

From Table 3, can be noticed that SSIM and FSIM present similar results for the four measures for both codecs. The SSIM and FSIM metrics present high Pearson and Spearman

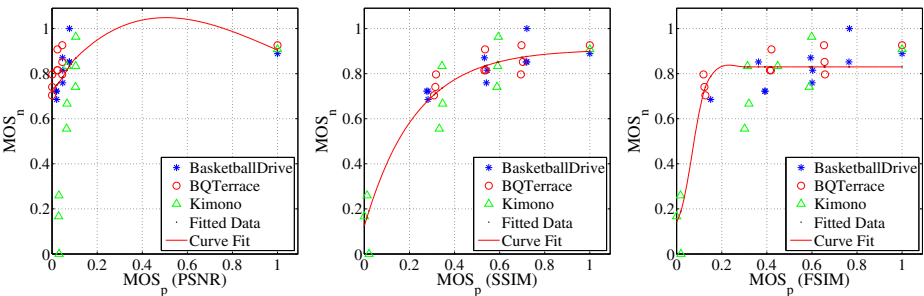
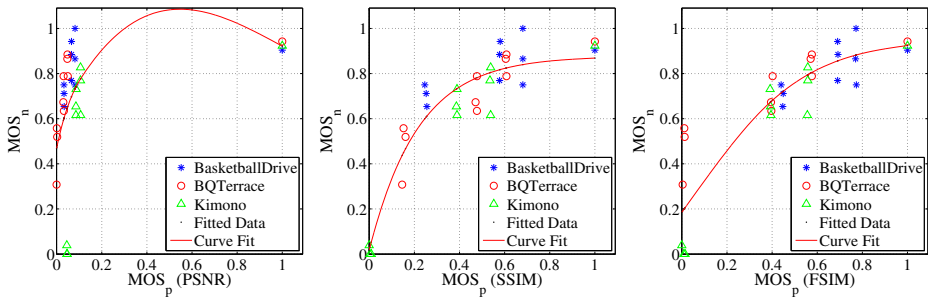


Fig. 6 HEVC fitting analysis for MOS vs PSNR, SSIM and FSIM metrics



**Fig. 7** VP9 fitting analysis for MOS vs PSNR, SSIM and FSIM metrics

correlations, while the PSNR presents low values. The outlier ratio is small and the RMSE presents similar values for all codecs and metrics.

The conclusions presented above are confirmed by the fitting analysis, of the MOS vs metrics, presented in Figs. 6 and 7 for the HEVC and the VP9 codecs respectively. These figures show that the FSIM and SSIM metrics follow the MOS variation and represent a good approximation. Moreover, the PSNR does not represent properly the MOS.

## 5 Conclusion

The goal of this study was to analyze and compare the subjective performance of the HEVC and VP9 codecs.

It was concluded that for the same values of the IP parameter, the HEVC codec result in lower bit-rates than the VP9 codec. As the bit rate value grows, both encoders provide a very similar perceptual quality that is also very similar to the obtained with the original videos. The statistical analyses confirms that the HEVC requires lower bit rates than the VP9 for the same subjective quality.

Hence, can be conclude that the VP9 codec did not succeed in providing the same encoding performance as the provided by the HEVC codec. As an advantage, the VP9 requires less computational effort than the HEVC. The differences on the computational requirements were not the goal of our study and are not reported.

The quality representation provided by three objective metrics (PSNR, SSIM and FSIM) were also studied and correlated with the subjective results. For both HEVC and VP9 codecs, the SSIM and FSIM metrics present a high correlation with the MOS, providing a valuable tool for objective quality evaluation. However, the commonly used PSNR presents a low correlation, and its usage should not be encouraged in this context.

**Acknowledgments** The authors are very grateful to the Instituto de Telecomunicações and Fundação para a Ciência e a Tecnologia (project UID/EEA/50008/2013), and to the Optics Center of Universidade da Beira Interior where this work has been conducted.

Appendix: settings for the encoders

Table 4 Selected setting for HM reference software encoder

Coding Options	Chosen Parameters
Encoder	HM 13.0
Profile	Main
Reference Frames	4
R/D Optimization	Enabled
Motion Estimation	TZ search
Search Range	64
Group of Pictures	8
Hierarchical Encoding	Enabled
Temporal Levels	4
Intra Period	1sec
Deblocking Filter	Enabled
Coding Unit (size / depth)	64/4
Transform Unit size (min / max)	4/32
Transform Skip	Enabled
Transform Skip Fast	Enabled
Hadamard ME	Enabled
Asymmetric Motion Partitioning (AMP)	Enabled
Fast Encoding	Enabled
Fast Merge Decision	Enabled
Sample Adaptive Offset (SAO)	Enabled
Rate Control	Disabled
Internal Bit Depth	8

Table 5 Selected setting for VP9 encoder

Chosen Parameters
--good --cpu-used=0
-- codec=vp9 -w < Width > -h < Height >
--i420 --psnr
--threads=0 --profile=0
--lag-in-frames= < FrameRate >
--min-q=< QP > --max-q=< QP >
--kf-min-dist=< IntraPeriod >
--kf-max-dist=< IntraPeriod >
--cq-level=20 --end-usage=0
--auto-alt-ref=1 --passes=2
--drop-frame=0 --static-thresh=0
--bias-pct=50
--minsection-pct=0 --maxsection-pct=2000
--arnr-maxframes=7 --arnr-strength=5
--arnr-type=3 --sharpness=0 --undershoot-pct=100

## References

1. Bankoski J, Bultje RS, Grange A, Gu Q, Han J, Koleszar J, Mukherjee D, Wilkins P, Xu Y (2013) Towards a next generation open-source video codec. Proc. SPIE 8666:866,606–866,606–13
2. Bultje R, Koleszar J Webm discussion. <https://groups.google.com/a/webmproject.org/>
3. Callet PL, Möller S, Perks A (eds) (2012) QUALINET White Paper on Definitions of Quality of Experience, Lausanne, Switzerland
4. De Simone F, Goldmann L, Lee JS, Ebrahimi T (2011) Towards high efficiency video coding: subjective evaluation of potential coding technologies. J Vis Commun Image Represent 22(8):734–748
5. Garcia R, Kalva H (2014) Subjective evaluation of HEVC in mobile environments. IEEE Trans Consum Electron 60(1):116–123
6. Grois D, Marpe D (2013) Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders. In: PCS 2013 30th picture coding symposium. San Jose, California
7. Hanhart P, Rerabek M, De Simone F, Ebrahimi T (2012) Subjective quality evaluation of the upcoming HEVC video compression standard. In: Proceedings of SPIE, applications of digital image processing, vol 8499. San Diego
8. Haskell B, Puri A, Netravali A (1996) Digital video: an introduction to MPEG-2. Digital multimedia standards series. Springer, US
9. Hollander M, Wolfe D (1999) Nonparametric statistical methods. Wiley
10. ISO/IEC 13818-2:2000 (2000) Information technology: generic coding of moving pictures and associated audio information, Part 2: Video. Standard, international organization for standardization
11. ISO/IEC 14496-2:1999 (2000) Information technology: coding of audio-visual objects, part 2: Visual. Standard, international organization for standardization
12. ITU-R Recommendation BT.500-12 (2009) Methodology for the subjective assessment of the quality of television pictures. Tech. rep. International Telecommunication Union
13. ITU-R Recommendation BT.2022 (2012) General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays. Tech. rep. International Telecommunication Union
14. ITU-T Recommendation H.263 (2005) Video coding for low bitrate communication. Tech. rep. International Telecommunication Union
15. ITU-T Recommendation H.264 (2012) Advanced video coding for generic audiovisual services. Tech. rep. International Telecommunication Union
16. ITU-T Recommendation P.910 (2008) Subjective video quality assessment methods for multimedia applications. Tech. rep. International Telecommunication Union
17. ITU-T Tutorial (2004) Objective perceptual assessment of video quality: full reference television. Tech. rep. International Telecommunication Union
18. Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. J Amer Stat Assoc 46(253):68–78
19. Mitchell J, Pennebaker W, Fogg C (2000) Mpeg video: compression standard. Kluwer Academic
20. MPEG (2013) Repository HEVC. <https://hevc.hhi.fraunhofer.de/svn/svn.HEVCSoftware/>
21. Ohm JR, Sullivan GJ, Schwarz H, Tan TK, Wiegand T (2012) Comparison of the coding efficiency of video coding standards – including high efficiency video coding (HEVC). IEEE Trans Circuits Syst Video Techn 22(12):1669–1684
22. Ostermann J, Tanimoto M (2012) MPEG video compression future. Springer
23. Rerabek M, Ebrahimi T (2014) Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments. In: SPIE optical engineering + applications. San Diego
24. Rouse D, Hemami S (2008) Understanding and simplifying the structural similarity metric. In: 15th IEEE international conference on image processing, 2008. ICIP 2008, pp 1188–1191
25. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611
26. Sullivan GJ, Ohm JR, Han W, Wiegand T (2012) Overview of the high efficiency video coding (hevc) standard. IEEE Trans Circuits Syst Video Techn 22(12):1649–1668
27. Sullivan GJ, Wiegand T (2005) Video compression - from concepts to the H.264/AVC standard. Proc IEEE 93(1):18–31
28. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

29. WebM WebM™: an open web media project - VP9 video codec summary. <http://www.webmproject.org/vp9/>. November, 20 of 2014
30. WebM WebM™: an open web media project, VP8 encode parameter guide, 2-pass best quality VBR encoding project. <http://www.webmproject.org/docs/encoder-parameters/>. June, 17 of 2013
31. Weerakkody R, Mrak M, Baroncini V, Ohm JR, Tan TK, Sullivan G (2014) Verification testing of hevc compression performance for uhd video. In: 2014 IEEE Global conference on signal and information processing (GlobalSIP), pp 1083–1087
32. Wiegand T, Sullivan GJ, Bjontegaard G, Luthra A (2003) Overview of the h.264/avc video coding standard. IEEE Trans Circ Syst Video Technol 13(7):560–576
33. WMA (2009) World medical association declaration of Helsinki - ethical principles for medical research involving human subjects
34. Zhang L, Zhang D, Mou X, Zhang D (2011) FSIM: a feature similarity index for image quality assessment. IEEE Trans Image Process 20(8):2378–2386



**Pedro Fernandes** was born in Penajóia, Portugal, in 1992. He received the B.S. degree in informatics engineering from the University of Beira Interior, Portugal, in 2014. His current research interests include codification in Real Holograms and Synthesized Holograms.



**Marco V. Bernardo** was born in Leomil, Portugal, in 1980. He received the B.S. degree in computer engineering from the University of Beira Interior (UBI), Portugal, and the M.Sc. degree in computer engineering, specializing in networks and multimedia, from the UBI, Portugal, in 2007 and 2009, respectively. He is currently pursuing the Ph.D. degree with the UBI, Portugal. His current research interests include image processing, image quality assessment, color quality and hyperspectral imaging.





**Antonio M. G. Pinheiro** (IEEE SM'15) was born in Covilha, Portugal, in 1965. He received the B.E. degree in electrical engineering from the I.S.T. University of Lisbon, Portugal, in 1988, and the Ph.D. degree in Electronic Systems Engineering from University of Essex in 2002. Since 1988 he is a lecture at U.B.I., Covilha, Portugal. Currently he is the head of the Optics Research Center of U.B.I. His current research interests are on image processing and computer vision domains, including Multimedia Quality, Multimedia Privacy, and also on Image Classification and Medical Image Analysis. He is a Portuguese committee member of ISO/IEC JTC 1/SC 29/WG 1 (JPEG). He has been Portuguese representative of the European COST Actions IC1003 Qualinet, and 292, and currently is Portuguese representative of IC1206, DE-ID - De-identification for privacy protection in multimedia content and BM1304, MYO-MRI - Applications of MR imaging and spectroscopy techniques in neuromuscular disease.



**Paulo Torrão Fiadeiro** is currently a full professor in the Department of Physics of the Faculty of Sciences at the University da Beira Interior (UBI) and a senior Applied Optics researcher in the Optics Center at UBI. He is also a collaborator in the Fibre Materials and Environmental Technologies (FibEnTech) research unit at the UBI. He holds a graduate degree in Electrotechnical Engineering Digital Systems and Computers by the Instituto Superior Técnico (IST) of the University of Lisbon (UL), Portugal (1986), a doctoral degree in Physics Optics by the Loughborough University of Technology (LUT), United Kingdom (1995), and an Aggregation degree in Physics by the UBI (2005). His main scientific domains and interests are Optics and Vision Sciences, namely digital holography, laser interference and diffraction, colour, hyperspectral imaging and visual acuity preferential looking.



**Manuela Pereira** received the 5-year BS degree in Mathematics and Computer Science in 1994 and the MS degree in Computational Mathematics in 1999, both from the University of Minho, Portugal. She received the PhD degree in Signal and Image Processing in 2004 from the University of Nice Sophia Antipolis, France. She is an Assistant Professor of Computer Science at the University of Beira Interior, Portugal. Her main research interests include: image/video coding, multiple description coding, joint source/channel coding, HDR, holographic and plenoptic image/video coding, Quality of Experience (QoE), QoE assessment and QoE modeling. She is the co-editor of 6 books and has authored or co-authored around 70 papers in international refereed journals and conferences. She served as a technical and program committee member for several IEEE journals and conferences. She is also a member of the editorial review board of several International Journals.