Study and evaluation of four video codecs –H.264, H.265, VP8 and VP9– for public transport entertaining systems

Francisco Fraile, Natalia Blasco, Ismael de Fez, Román Belda, Pau Arce and Juan Carlos Guerri

Instituto de Telecomunicaciones y Aplicaciones Multimedia,
Universitat Politècnica de València,
8G Building - access D - Camino de Vera s/n - 46022 Valencia (Spain)
Corresponding author: jcguerri@iteam.upv.es

Abstract

This paper presents an evaluation of four video codecs, H.264, HEVC, VP8 and VP9. The evaluation focuses on a video encoding platform for an entertainment system for vehicles. The paper describes the differences between the four codecs as well as the main methodologies and metrics used to evaluate the performance of video codecs. Results include both objective metrics and subjective metrics. The objective evaluation results show that although VP9 can achieve the best trade-off between PSNR and bitrate, only the H.265 and H.264 implementations meet the real time encoding requirement with sufficient video quality. Finally, the subjective evaluation –conducted with simulation software that emulates viewing conditions inside a vehicle– shows no evidence to support the choice of H.265 over H.264.

Keywords: video coding, H.264, HEVC, VP8, VP9, Quality of Experience (QoE)

1. Introduction

Public transport services like coaches, trains or boats, need to provide their customers with added value services in order to differentiate their service from the competition. In this context, entertainment systems –gathering services like Internet access, Video on Demand and live television and radio– help passengers spend the time of the journey more comfortably and represent a great asset in any public transport vehicle.

In these entertainment systems, the requirements for video encoders are quite challenging and must be taken into account when choosing the right video encoding technology. One differentiating feature of an on board entertainment system is to allow access from user devices like laptops, tablets and mobile phones to the platform. Passengers are used to interact with their devices during the journey and therefore they are eager to access the platform through them. This constrains video encoding technology and formats to those compatible with consumer electronic devices. In addition to this, video distribution needs to be wireless, normally through a Wi-Fi network installed in the vehicle. In order to ensure a good Quality of Service, the video encoding technology needs to provide a good trade-off between bitrate and video quality. Furthermore, offering live television means that the video needs to be encoded in real time.

Taking these requirements into account, this paper shows in section 2 the state of the art of video encoders for consumer electronics. Later, section 3 describes the main metrics used to evaluate the trade off between rate and distortion of video encoding technologies. Later, some test sequences encoded with different video encoding technologies are evaluated and the results are analyzed, in order to determine which is the best technology for the case under study.

2. State of the art

2.1 Introduction

Currently, there are several video encoders leading the market. Specifically, on one hand there is the popular and proprietary MPEG-4 H.264 encoder and its evolution H.265-HEVC and on the other hand the free encoders developed by Google VP8 and its successor VP9.

Video quality metrics play an important role when determining which encoders and what encoding parameters achieve an optimum performance in a given scenario.

Together, these four encoders take up for most of commercial video applications and consequently they will be analyzed in this study. Section 2.2 provides a brief overview of these encoders.

Later on, the paper evaluates the impact that the encoding process has on the video quality using different methodologies. In this context, video quality metrics play an important role when determining which encoders and what encoding parameters achieve an optimum performance in a given scenario. Section 2.3 presents some of these metrics. Finally, Section 2.4 describes the software used to conduct the measurements presented in this article.

2.2 Video encoders

2.2.1 H.264

The International Telecommunication Union –Video Coding Experts Group (ITU-VCEG)– and the Moving Picture Experts Group (MPEG) formed a Joint Video Team (JVT) in 2001 in order to develop the recommendation H.264-Advance Video Coding (AVC), or MPEG-4 part 10 [1]. The aim was to improve the performance of previous standards while keeping the syntax simple. H.264 achieves greater encoding efficiency providing good picture quality with reduced bitrate.

The main innovation included in the recommendation is the definition of a two layer model: the Video Coding Layer (VCL), representing the encoded video data, and the Network Abstraction Layer (NAL), providing the formats for payload and header information for the transport of encoded video on different technologies or the storage in different media formats [1]. With this, the H.264 recommendation covers several video applications, like digital storage, television broadcasting, video transmission over IP networks or videoconference, among others.

2.2.2 H.265

The JVT recommendation High Efficiency Video Coding (HEVC-H.265) [2] was developed in 2010 and is the latest video coding recommendation from the JVT group at the time of writing. H.265 implementations are expected to improve 50% the video compression of H.264 with the same picture quality, at the expense of increasing computational requirements. HEVC is nonetheless very well suited for online video distribution, since it reduces the bandwidth requirements and the video is encoded offline. This is why H.265 has a strong backing from the Internet media industry. In addition to Internet streaming, H.265 covers a wide range of video applications, just as H.264.

2.2.3 VP8

Google acquired On2 Technologies – developer of the VP7 encoder- in 2010. That very same year Google released the source code of its evolution, VP8 [3]. VP8 is based on the decomposition of frames into square sub-blocks of pixels, the prediction of such sub-blocks using previously constructed blocks, and the adjustment of such predictions (as well as the synthesis of unpredicted blocks) using a Discrete Cosine Transform (DCT). VP8 works only with 8-bit YUV 4:2:0 image formats. Later, Google developed the open source video container WebM providing an open source alternative for video applications. VP8 and WebM are included in the specifications of HTML5 video and are supported in applications like YouTube and browsers like Chrome, Opera, Mozilla and others.

2.2.4 VP9

VP9 [4], released in 2011, is the evolution of VP8. VP9 was developed to improve 50% the video compression of VP8 with the same picture quality, as well as improving the encoding efficiency. A large part of the encoding efficiency improvements in VP9 come from the introduction of larger prediction block sizes, super-blocks of size up to 64x64. This encoder is supported in applications such as Microsoft Edge, Google Chrome, Mozilla Firefox, VLC and FFmpeg. Moreover, VP9 is particularly suited for 4K video streaming applications, as it can reduce 50% the data throughput.

2.3 Parameters for video coding evaluation

There are two different families of tests for video quality assessment, namely subjective tests – using test subjects to detect artifacts in encoded video sequences– and objective tests –using algorithms that estimate the quality of the encoded video.

Regarding the former, the ITU-R BT.500 recommendation describes several methods to standardize subjective tests, containing procedures and requirements to choose and configure adequate displays, select test subjects or determining optimum test and reference video sequences. Thus, ITU-R BT.500 provides a methodology to conduct subjective evaluations in a formal way.

The most representative metric concerning subjective tests is the Mean Opinion Score (MOS). The MOS is a subjective metric where test subjects rate the perceived quality on different video sequences in a scale from 1 to 5, being 5 the best rating. The MOS is then generated as the average over a set of subjective evaluations provided by the test audience. The most basic method to obtain the MOS is known as Single-Stimulus (SS). In this method, test subjects watch the encoded video once and provide a rating from 1 to 5. Another method present in ITU-R BT.500 is referred to as the Double-Stimulus Continuous Quality Scale (DSCQS). With this method, test subjects watch both the reference sequence and the test sequence twice, but without knowing which ones are the reference sequences. Later, test subject must provide a rating to each video sequence. The results of the DSCQS are represented by the Differential Mean Opinion Score (DMOS), which is a metric calculated as the difference between the average MOS rating of the reference sequence and the average MOS rating of the test sequence. This way, a negative DMOS implies that the test sequence was perceived as having better quality than the reference sequence whereas a positive DMOS represents quality loss.

Alternatively, there are several objective video quality metrics in order to assess the quality of video sequences. The most basic and frequently used metric is the Peak Signal to Noise Ratio (PSNR). The PSNR computes the differences between the pixels of each frame in the reference sequence and the corresponding pixels (in time and space) of the test sequence. Although the PSNR is a good metric to evaluate the quality of the encoding process, the PSNR does not provide a good correlation with subjective tests in all circumstances. For this reason, the PSNR alone is not sufficient to assess encoded video quality.

The PSNR belongs to one of the two different categories of objective quality measurements based on references, named noise based measurements. The other category – based on perception models– uses human vision system models in order to determine the perceptual contrast of reference and test sequences. This perceptual evaluation of contrast takes into account perceptible changes of measurable parameters like contrast. Later, other perceptual characteristics like the relationship between contrast and luminance between the two sequences are processed, as well as other properties of the behavior of the human eye that can mask differences between both images. This way, metrics compute the differences between perceptible changes between the reference and the test sequence, instead of just differences between pixels. Thus, the difference in the perceptual contrast is used to perform video quality measurements based on perception models. Provided that the vision model is accurate enough, these metrics can achieve a high degree of correlation with subjective tests.

Within this category, the Structural Similarity Index (SSIM) is an alternative method to PSNR to predict subjective video quality based on a perception model. This is a closer approach to subjective metrics since the SSIM accounts for characteristics of the human vision system that can give little importance to changes in pixel information, like changes in the luminance.

The Picture Quality Rating (PQR), the predictive DMOS and the Attention weighted DMOS (ADMOS) are other metrics based on perceptual models that use more sophisticated human vision system models. These methods exhibit higher correlation with subjective tests than the PSNR. The PQR converts perceptual contrast differences between the reference and the test video sequences in a value that

Subjective tests use test subjects to detect artifacts in encoded video sequences and objective tests use algorithms that estimate the quality of the encoded video.

represents the ability of the spectator to perceive those differences between videos, in units called Just Noticeable Differences (JNDs). This way, the PQR evaluates whether spectators are able to tell the differences between the test video and the reference video. For this reason, the PQR is very well suited for high quality applications like High Definition TV (HDTV).

On the other hand, the predictive DMOS and the ADMOS evaluate the impairments between the test video and the reference video as perceived by spectators. These impairments are quantified in the same scale from 1 to 5 that is used in subjective test surveys, as described above. These methods are suited to evaluate wide ranges of video qualities (unlike PQR which as stated is only suitable for high definition video). In the scale of DMOS and ADMOS, values between 0 and 20 represent an excellent subjective quality, values between 20 and 40 represent an acceptable quality, whereas values in the range between 40 and 100 represent a bad quality.

The only difference between the predictive DMOS and the ADMOS is that the ADMOS model implements a human cognitive component that takes into consideration the area of the scene where spectators will focus their attention. Therefore, impairments around this area will have a much greater impact in the ADMOS, whereas impairments far from the area will have little or no impact. With this, the ADMOS provides a better correlation with subjective tests and is a reliable alternative as shown in [5].

2.4 Software for video evaluation

2.4.1 FFmpeg

FFmpeg is an open source software project that contains programs and libraries to handle both audio and video flows. FFmpeg allows to convert audio and video into a wide range of formats. It can convert between arbitrary sample rates and resize video on the fly with a high quality polyphase filter. FFmpeg supports several video, audio, image and subtitles formats, among them: MPEG-4, H.264, H.265, MP3, AAC, WMA, etc.

2.4.2 PQA SW

The evaluation of the ADMOS of the different video sequences has been conducted with the Tektronix PQA SW. This software allows to estimate with great accuracy the DMOS of subjective tests conducted with human subjects. Unlike subjective tests, the PQA SW software provides an estimation of the DMOS per each frame in the video sequence, thus giving a better understanding of which frames or sequences of the video test represent

Figure 1. Block diagram of PQA SW used for ADMOS evaluation.

a greater challenge to the encoder. The PQA SW implements a user interface to configure the human vision system model in a visual manner, from a block diagram. In this study, the different blocks have been configured to simulate viewing conditions in an on board entertainment system. More specifically, the configuration implements the block diagram shown in Figure 1.

Each of the blocks represents a configuration element of the human vision system model used by the software. For the tests, the display model is configured as an LCD type display with HD resolution (1920x1080 pixels). This simulates a display such as a smartphone or a tablet. As for the environment conditions, the viewing distance has been set to 3 times the height of the display and the ambient luminance has been set to 0.25 cd/m^2, which corresponds to a moderate luminance, as can be expected inside a vehicle. Regarding the perceptual model, the default configuration simulating the vision model of an average spectator has been used. Finally, the attention model applies a 66% attention probability to movement, a 100% weighting and a 100% attention probability to the differences between the test and reference sequences.

3. Methodology

The methodology consists of two main elements: a qualitative evaluation of the encoders to analyze their performance and a subjective evaluation to analyze the impact on the subjective quality. The qualitative evaluation uses the PSNR to determine the distortion produced in the video sequence by the four encoders under study at different encoding rates, while the subjective evaluation analyzes the impact on the subjective quality, represented by the ADMOS. In addition to the PSNR, other important metrics in the evaluation of the performance of video encoders are assessed, like the encoding time and the encoding speed.

In order to provide a first approach towards a subjective study, Table 1 presents a relationship between the PSNR and the MOS.

PSNR (dB)	MOS
>37	5 (Excellent)
31-37	4 (Good)
25-31	3 (Fair)
20-25	2 (Poor)
<20	1 (Bad)

Table 1. PSNR to MOS conversion [6].

The qualitative evaluation uses the FFmpeg encoding library, version 2.9. Following versions of video encoders have been used: x264-snapshot-20141218-2245-stable [7], x265 –obtained from Mercurial on July 2015 [8]– and libvpx-1.4.0 [9].

The EvalVid framework version 2.7 has been used in order to compare the quality of the different encoded videos. Moreover, two different video sequences have been selected to conduct the objective evaluation. The YUV formatted videos have been obtained from [10]. The length of the videos is 10 seconds. The first, *big_buck_bunny*, is encoded at 1080p and has 14 315 frames, while the second video, *crowd_run*, is encoded in 2K format and has 500 frames. Figure 2 show snapshots of these videos.

a) big_buck_bunny

b) crowd_run

Figure 2. Videos used for evaluation.

On the other hand, the subjective study consists of the evaluation of the ADMOS over a reference video of 30 seconds length (shown in Figure 3). This reference video has been obtained concatenating 3 different sequences (news, entertainment programs, sports, etc.) from real satellite television channels. It is expected that this sequence provides more representative results of the encoded video quality.

Finally, it is important to point out that both the qualitative and the subjective measurements have been performed using a Dell PowerEdge T110-2 with an Intel Xeon

Figure 3. Frames of the test sequence used.

processor at 3.5GHz and 32 GB RAM. It is important to include the specifications of the hardware because they can have an impact in the evaluation of the encoding time and the encoding speed. Although a more powerful hardware platform could provide higher encoding speeds, the conclusions regarding the comparison of the different encoders would remain the same.

4. Evaluation

4.1 Objective evaluation

4.1.1 Evaluation of the PSNR

This study evaluates the PSNR obtained by the codecs under study in a wide range of bit rates, over 2 test sequences. The evaluation of these two videos is shown below.

Figure 4 shows the evaluation of the PSNR of the video *big_buck_bunny*. Figure 5 shows the same results but in the range from 0 to 2 Mbps, which are the typical range of bitrates used for coding. Moreover, as shown in the figure, there are different areas (each one with a different color) which indicate the MOS level depending on the PSNR (by using the relation shown in Table 1).

First, it is important to indicate that the simulations have been performed varying the CRF which in turn changes the bitrate. However, the relationship between CRF is not linear and, although all codecs have been evaluated in a CRF range between 4 and 50, this does not produce the same range of bitrates, as the relationship between CRF and bitrate depends on the codec. For this reason, not all codecs cover the same range of bitrates in the x axis (for instance, in Figure 4 the bitrate of H.265 is in the range between 17 kbps and 19 Mbps, while the range of VP8 is between 1 Mbps and 14 Mbps).

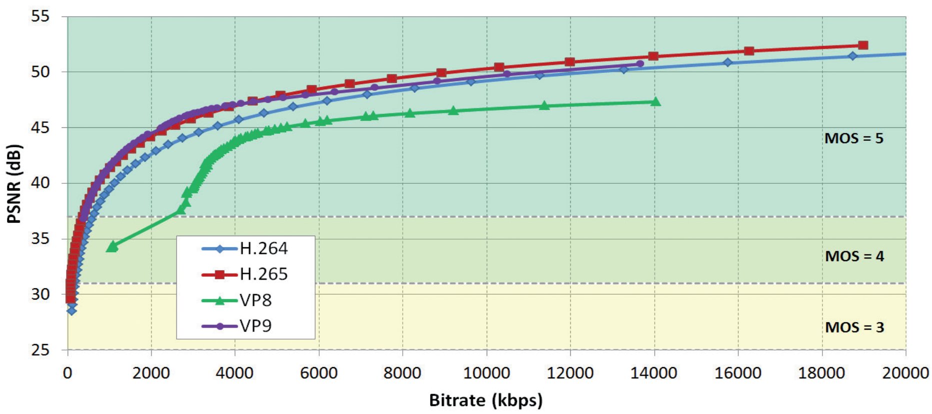


Figure 4. PSNR evaluation with video *big_buck_bunny*.

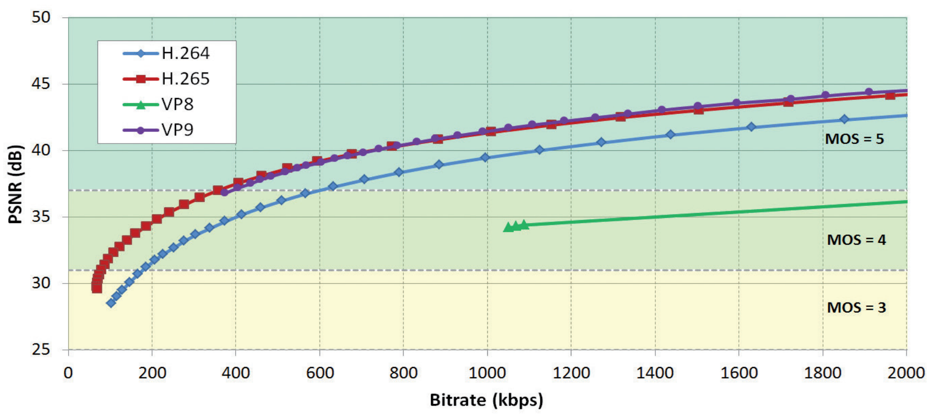


Figure 5. PSNR evaluation with video *big_buck_bunny* (detail).

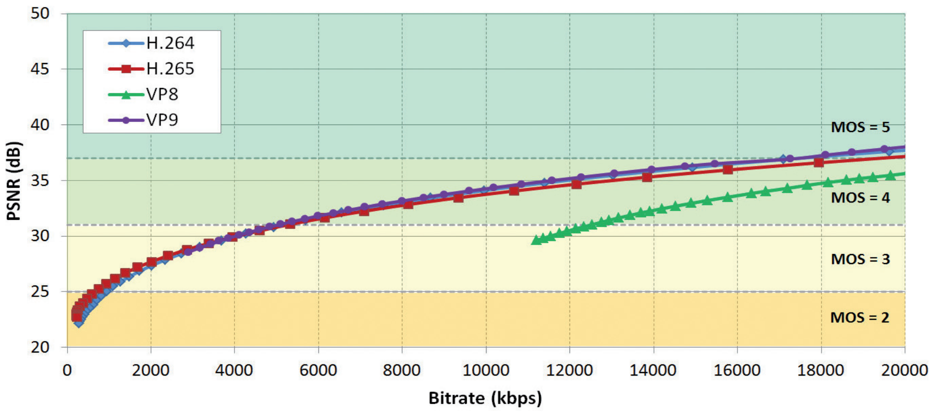


Figure 6. PSNR evaluation with video *crowd_run*.

As shown in the figure, in general, the H.264, H.265 and VP9 encoders perform better than VP8. For instance, to achieve an MOS = 5 (PSNR > 37) H.265 and H.264 need a bitrate of about 400 kbps, while the bitrate needed by H.264 is higher –600 kbps. VP8 on the other hand needs a considerably higher bitrate –2.7 Mbps. As seen in the figure, the behavior of H.265 and VP9 is practically the same in terms of PSNR, especially at lower bitrates. H.264 shows slightly worst results. Notice that the PSNR increases more rapidly at lower bitrates than at higher bitrates. This means that an increment of the bitrate at low bitrates results in a better improvement of the video quality than at higher bitrates, where there are fewer differences between the reference and the test sequences.

Figure 6 shows the evaluation of the bitrate for video 2. In this case, all codecs provide lower PSNR values than with video 1. The fact that the same codec needs a much higher bitrate to achieve the same PSNR proves that this video is more challenging for the encoders. In this scenario, VP9 provides the best results and H.264 brings similar results to H.265, even slightly better. VP8 shows the worst performance. With this video, to achieve an MOS = 5 VP9 needs a bitrate higher than 18 Mbps, H.264 a bitrate higher than 19 Mbps, H.265 higher than 20 Mbps and VP8 a bitrate higher than 23 Mbps.

4.1.2 Evaluation of the encoding time and speed

The second study evaluates one of the key parameters when evaluating a codec: the encoding time, and a related parameter like the encoding speed. This is a critical parameter since a high encoding time can mean the infeasibility of using a certain encoder depending on the execution environment. For instance, in a real time video coding environment, the encoding time should be minimized in order not to get worse the Quality of Experience of the user. In contrast, in a video-on-demand environment without temporal limitations, this parameter is not so critical.

Figure 7 and Figure 8 show the evaluation of the encoding time depending on the PSNR for videos *big_buck_bunny* and *crowd_run* respectively. As expected, getting a better PSNR implies a considerable increase of the encoding time. In fact, we can appreciate that the encoding time has an exponential growth regarding the PSNR. On the other hand, when comparing the codecs, the most striking feature is the big difference regarding the encoding time of VP9 compared to the rest of encoders. On the other hand, H.264 is the less exigent codec regarding the encoding time, whereas H.265 and VP8 provide higher values for the entire range of PSNR. Comparing these two codecs, we realize that the

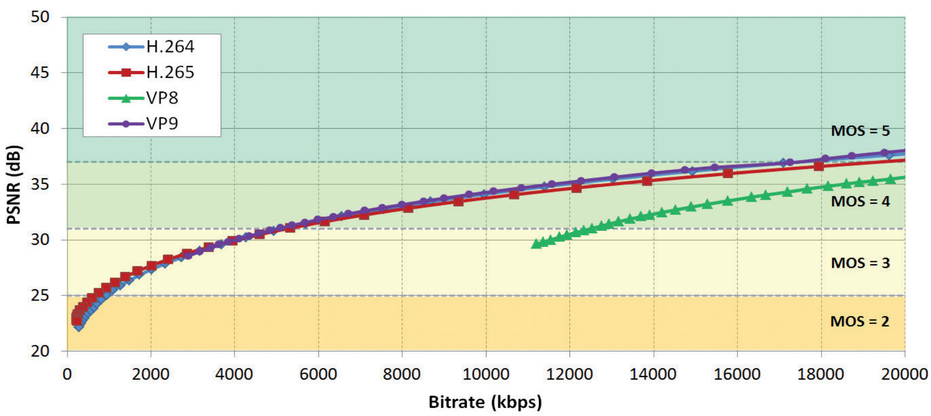
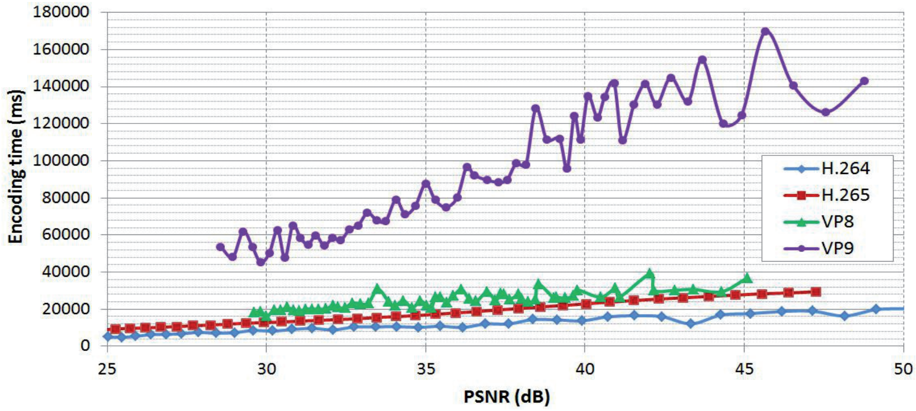


Figure 7. Encoding time evaluation with video *big_buck_bunny*.



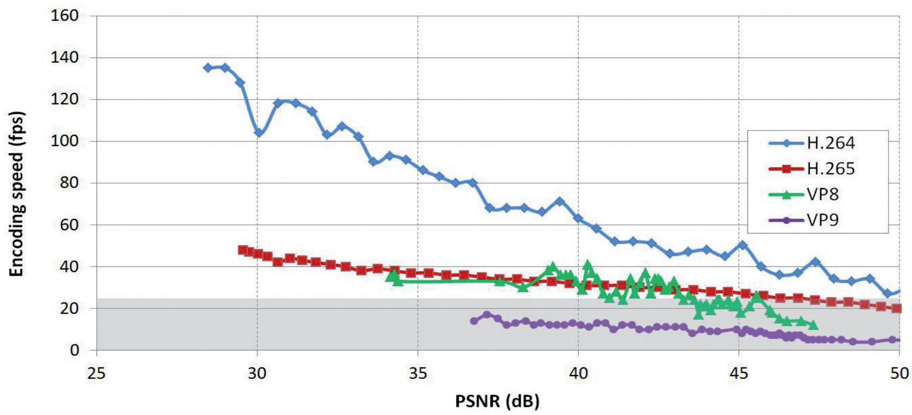
■ **Figure 8.** Encoding time evaluation with video *crowd_run*.

behavior of VP8 is much more irregular regarding the encoding time and, although rather similar, VP8 provides, on average, higher encoding times than H.265. The same conclusions arise in both videos. The only difference is that the encoding time of the video *crowd_run* is higher (note the scale of axis *y*). This makes sense considering the results obtained from the PSNR evaluation (Figure 4 and Figure 6), which showed that the first video provides higher values of PSNR than the second one.

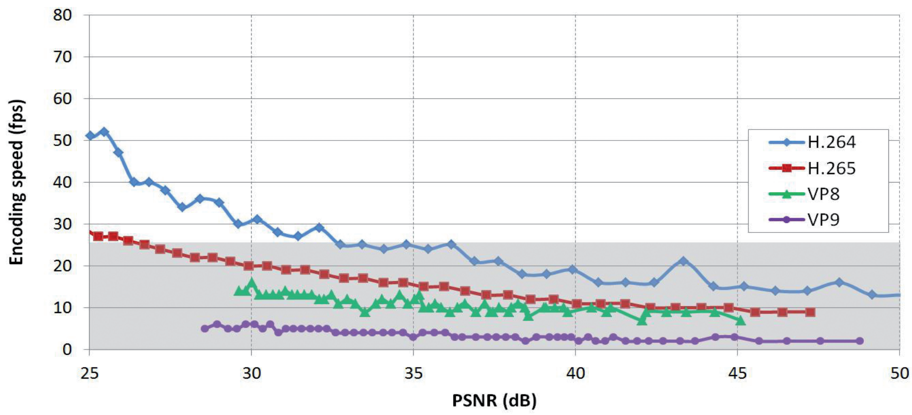
In order to compare the codecs in a more qualitative way, we are going to analyze in the video *big_buck_bunny* the encoding time needed by each codec to provide an 'Excellent' MOS, that is, a PSNR higher than 37 dB. H.264 requires an encoding time of 3.6 s, H.264 of 7.4 s, VP8 of 7.6 s and finally VP9 requires a encoding time of 15.3 s (325% higher than H.264).

The encoding time is related to the encoding speed, which is used to indicate the number of frames per second a codec is able to encode. Figure 9 presents the speed time evaluation for the video *big_buck_bunny*. In this case, an increase of the PSNR implies coding less frames per second, as the figure depicts. In the figure we can see that H.264 is able to encode more than 100 fps with levels of PSNR lower than 33 dB. To accomplish the condition of MOS = 5, H.264 can encode with a speed until 68 fps. On the other hand, H.265 and VP8 provide a similar performance (but far below H.264) encoding at 33 fps to get an MOS = 5. Finally, the encoding speed of VP9 is really low, and in this study does not even exceed 25 fps. Note that in the figure there is a grey area for values lower than 25 fps, since this value is used by the TV European system PAL (Phase Alternating Line). Therefore, in real time coding environments, that value represents a minimum encoding speed. Taking this into consideration, according to the figure, VP9 is not able to encode with an encoding speed higher than that limit, so this codec would not be appropriate to encode this video in a real time coding environment.

The evaluation of the speed time of the video *crowd_run* depicted in Figure 10 shows results in line with the previous study. That is, the encoding speed is much lower than in the video *big_buck_bunny*, and H.264 is the codec with the highest encoding speed, in contrast to VP9, which provides the lowest encoding speed. In this case, it is not possible to achieve a level of MOS = 5 in real time encoding environments, since for all codecs the encoding speed required to have a minimum PSNR of 37 dB is lower than 25 fps. To achieve a 'Good' subjective quality (that is, a PSNR higher than 31 dB) in a real time encoding environment we only could use H.264 according to the results.



■ **Figure 9.** Speed time evaluation with video *big_buck_bunny*.



■ **Figure 10.** Speed time evaluation with video *crowd_run*.

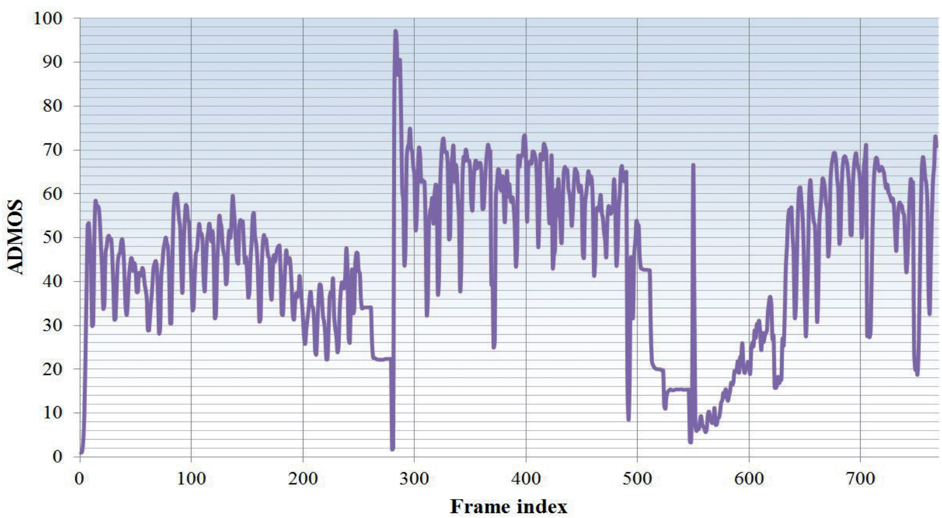
4.2 Subjective evaluation

This study used the PQA software to evaluate the subjective video quality. The video sequence used is the test sequence obtained from the satellite (video 3), encoded with H.264, H.265 and VP9 in two different resolutions –768x432 and 1280x720– and with different encoding rates. The objective of this study is to analyze in detail the behavior of the different codecs with a representative video sequence.

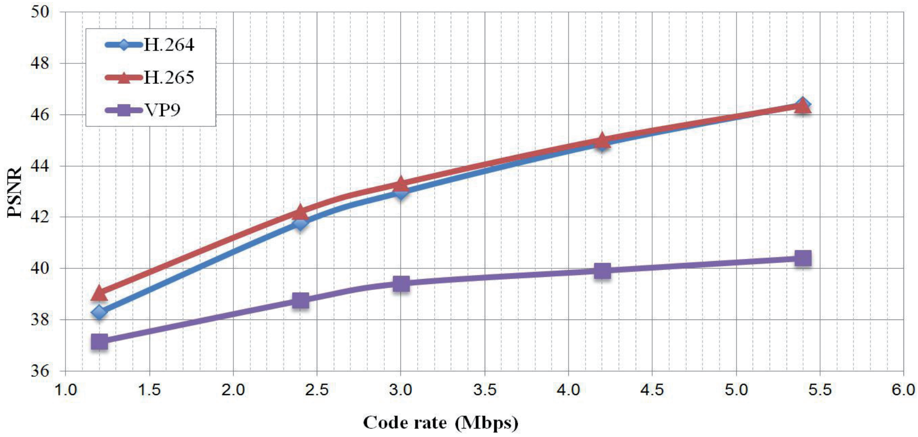
Figure 11 shows the ADMOS against the frame index for the 768x432 video encoded at 1.2 Mbps with VP9. The results show three clearly differentiated regions (frame 0 to 280, 280 to 493 and 493 to 755) corresponding to each of the video segments composing the test sequence. In the first segment, the ADMOS is more stable and lower than for the other two sequences. This highlights that this sequence is less challenging for the encoders and that there are less noticeable differences between the test and the reference videos. The picture changes all of a sudden between sequence 1 and 2, producing a peak in the ADMOS. Later, the ADMOS of sequence two is the highest of the three sequences. Finally, sequence 3 starts with a transition from white and, for this reason, there is no such peak as that between sequences 1 and 2. This way, the ADMOS for the third sequence increases with the complexity of the image.

Regarding the 768x432 resolution, the PSNR of the sequence encoded with VP9 at 1.2 Mbps is the lowest among all test sequences. Thus, the Minkowski index obtained from this sequence is used as the worst case training for all the test sequences of this resolution. Figure 12 shows the PSNR of the different encoders at the encoding rates selected for this study. The figure shows that the PSNR of H.264 and H.265 is slightly higher than the PSNR of VP9. The PSNR of the two MPEG encoders is very similar for all encoding rates. Moreover, the differences with VP9 are relatively low, especially at low encoding rates. In all cases, the PSNR is higher than 37 dB. Thus, the MOS corresponding to all the cases under study is 5 using the reference in Table 1.

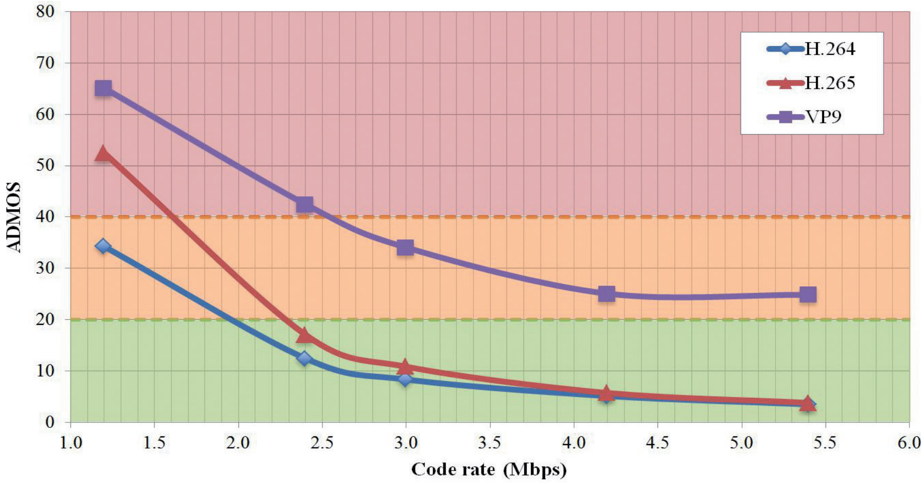
Figure 13 shows the average ADMOS for the test sequences obtained with the PQA software. The results



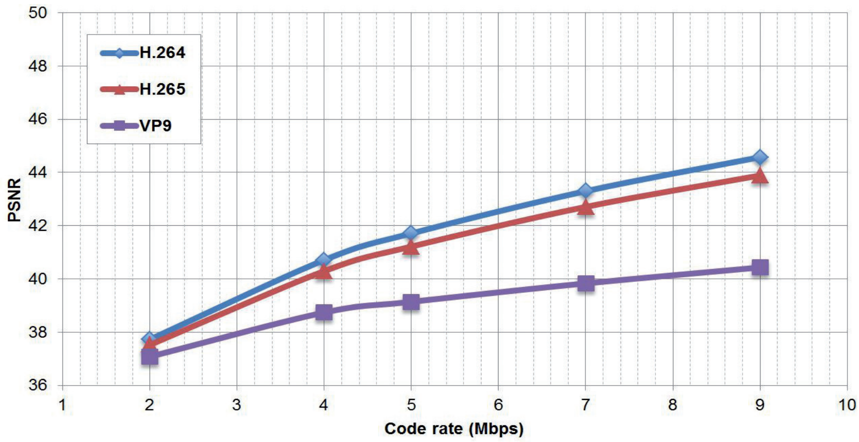
■ **Figure 11.** Detail of the ADMOS trace for resolution 768x432 using VP9 with video *video_ref*.



■ **Figure 12.** PSNR evaluation for resolution 768x432 with video *video_ref*.

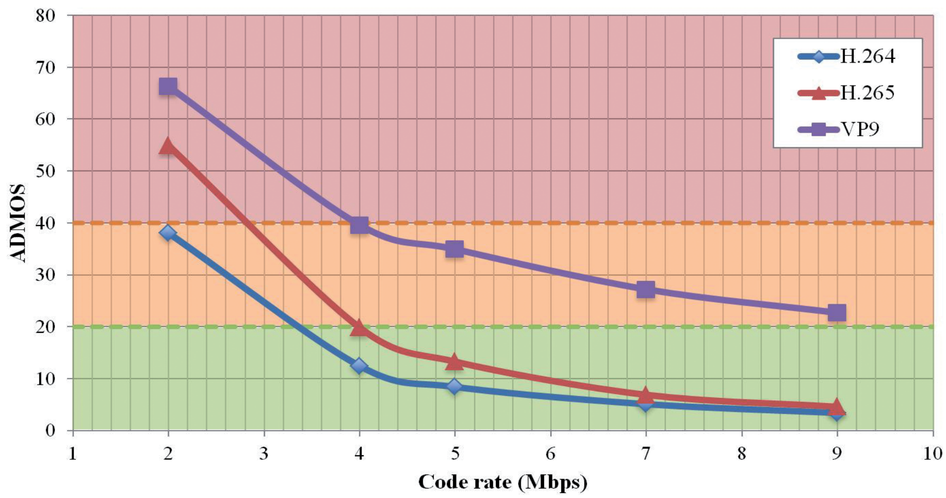


■ **Figure 13.** ADMOS evaluation for resolution 768x432 with video *video_ref*.



■ **Figure 14.** PSNR evaluation for resolution 1280x720 with video *video_ref*.

are, in general, consistent with the PSNR, although the differences in the ADMOS are much more noticeable allowing a thorough analysis of the subjective quality provided by each encoder at the different encoding rates. The graph shows three different regions according to the ADMOS subjective quality (0-20: good, 20-40: acceptable, 40-100: bad). The results show that the codec H.264 provides a good video quality with 1.2 Mbps encoding rate. On the other hand, H.265 and VP9 provide a bad quality for the same encoding rate. At this encoding rate, the H.265 provides the highest PSNR. However, the H.264 provides the best ADMOS. As the encoding rates increases, the subjective evaluation of H.265 and H.264 becomes more similar and the results are practically equivalent for both encoders. At 2.4 Mbps, the subjective quality of H.264 and H.265 is good,



▪ **Figure 15.** ADMOS evaluation for resolution 1280x720 with video *video_ref.*

whereas the ADMOS of VP9 is still higher than 40. For the remaining encoding rates, the subjective quality of H.264 and H.265 is good, whereas the subjective quality of VP9 is acceptable.

Figure 14 shows the average PSNR achieved by the three encoders under study at different encoding rates. The results are very similar for H.264 and H.265, the former providing a slightly higher PSNR, especially at high encoding rates. The PSNR of VP9 is slightly lower. In all cases, the PSNR is greater than 37 which, according to Table 1, corresponds to an MOS = 5.

On the other hand, Figure 15 shows the mean ADMOS for resolution 1280x720. The results show higher differences between the three codecs, especially at lower encoding rates. The H.264 provides acceptable quality (ADMOS lower than 40) with an encoding rate of 2 Mbps. For the same encoding rate, H.265 and VP9 provide a bad subjective quality. At higher encoding rates, the ADMOS of H.264 and H.265 is practically the same and in all cases lower than VP9. This is consistent with the PSNR results.

5. Conclusion

This paper has analyzed the video encoders most widely used in the market: H.264, H.265, VP8 and VP9, together with the main objective and subjective metrics used to evaluate encoded video quality. Regarding the objective evaluation, in terms of PSNR, VP9 provides the best results of the four codecs under study. The performance of H.265 is very similar, sometimes better than VP9, especially at low encoding rates, which are representative of the case under study. As a general conclusion, H.264 performs worse than H.265 and VP9 but better than VP8.

In terms of encoding time, the objective evaluation shows that the encoding time of VP9 is extremely long compared to the other codecs. H.264 shows on the other hand the shortest encoding times in all study cases. Moreover, in general, the encoding time of H.265 is lower than VP8. The latter varies with PSNR significantly. In this sense, the study has shown that the only two codecs capable of encoding in real time with a MOS ≥ 4 (at least 25 fps) are H.264 and H.265.

This way, the objective study proves that VP8 has the worst performance of all codecs. Although VP9 brings the best results, the differences with H.265 and H.264 are small. However the encoding time of VP9 is much higher than the encoding time of H.264 and H.265. This is relevant in a real time encoding scenario like the study case. With respect to H.265 and H.264, the objective evaluation shows better results for H.265 although the differences are not very significant. In this sense, the properties of the video (like the complexity of the scene) as well as the encoding parameters can explain the differences between different comparisons of these two codecs.

As for the subjective evaluation, H.264 provides the lowest mean ADMOS. These results are consistent with the PSNR obtained for the same video sequence. In comparison with the PSNR, the ADMOS shows greater differences between the different codecs. Therefore conclusions obtained from their analysis may be different and it is necessary to take both into account in order to perform an exhaustive analysis of the performance of different encoders in a given scenario.

Acknowledgments

This work is supported by the Ministerio de Economía y Competitividad of the Government of Spain under project "Plataforma avanzada de conectividad en movilidad (CONECTMOV) (IDI-20150126)".

References

[1] International Telecommunication Union, "Advanced video coding for generic audiovisual services, Recommendation ITU-T H.264," Mar. 2003.

[2] International Telecommunication Union, "High Efficiency Video Coding, Recommendation ITU-T H.265," Apr. 2015.

[3] J. Bankoski, J. Koleszar, L. Quillio, J. Salonen, P. Wilkins and Y. Xu, "VP8 Data Format and Decoding Guide," RFC, vol. 6386, Nov. 2011.

[4] A. Grange and H. Alvestrand, "A VP9 bitstream overview," Internet-draft, Feb. 2013.

[5] Objective Measurements and Subjective Assessments, Application Note. PQA White Paper, Tektronix.

[6] J. R. Ohm, "Multimedia communication technology: representation, transmission and identification of multimedia signals," Springer Berlin Heidelberg, New York, 2004.

[7] Encoder version x264-snapshot-20141218-2245-stable, available online at: https://ftp.videolan.org/x264/ snapshots/.

[8] Encoder version x265-1.9, available online: https://bitbucket.org/multicoreware/x265/wiki/Home.

[9] Webmproject.org, encoder version libvpx-1.4.0, available: http://www.webmproject.org/code/.

[10] Xiph.org video test media, available online: https://media.xiph.org/video/derf/.

[11] T. K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J.-R. Ohm, and G. Sullivan, "Video quality evaluation methodology and verification testing of HEVC compression performance," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, 2016.

Biographies

Francisco Fraile obtained a degree in Telecommunication Engineering from the Universitat Politècnica de València (UPV) an M. S. Degree in Microwave Engineering from the University of Gävle in 2004, a M.S. degree in Telematics from UPV in 2009 and his PhD in telecommunications in 2013. From 2004 to 2010, he worked as a Research Engineer for the Swedish company Interactive TV Arena. In 2006, he joined the Multimedia Communications research group (COMM) of the Institute of Telecommunications and Multimedia Applications (iTEAM), UPV.

Natalia Blasco obtained a degree in Telecommunication Engineering from the Universitat Politècnica de València (UPV) in 2016. Currently, she is a Researcher at the Multimedia Communications research group (COMM) of the Institute of Telecommunications and Multimedia Applications (iTEAM), UPV. Her areas of interest are video coding and multimedia QoS.

Ismael de Fez received the Telecommunications Engineering degree and the M.S. degree in telematics from the Universitat Politècnica de València (UPV), Valencia, Spain, in 2007 and 2010, respectively. In 2014 he obtained his PhD in Telecommunications from the UPV. His doctoral thesis was awarded by the UPV. Currently, he is a Researcher at the Multimedia Communications research group (COMM) of the Institute of Telecommunications and Multimedia Applications (iTEAM), UPV. His areas of interest are file transmission over unidirectional environments and file encoding.

Román Belda received the Computer Science degree in 2004 and the M.S. degree in telematics in 2013 from the Universitat Politècnica de València (UPV), Valencia, Spain. He currently works as a Researcher at the Multimedia Communications research group (COMM) of the iTEAM Institute, where he is working toward the PhD degree. His areas of interest are mobile applications and multimedia transmission protocols.

Pau Arce received his Telecommunications Engineering degree and the M.S. in Telematics from the Universitat Politècnica de València (UPV), Spain, in 2005 and 2007 respectively. In 2014 he obtained his PhD in Telecommunications from the UPV. Currently he works as a researcher at the Institute of Telecommunications and Multimedia Applications (iTEAM). His research interests include multimedia QoS, routing on wireless ad hoc networks and performance evaluation of computer systems.

Juan Carlos Guerri was born in Valencia. He received his M.S. and Ph. D. (Dr. Ing.) degrees, both in telecommunication engineering, from the Universitat Politècnica de València (UPV), in 1993 and 1997, respectively. He is a professor in the E.T.S. Telecommunications Engineering at the Universitat Politècnica de València, where he leads the Multimedia Communications research group (COMM) of the iTEAM Institute. He is currently involved in research and development projects for the application of multimedia to industry, medicine, education, and communications.