

Performance analysis of VP8 image and video compression based on subjective evaluations

Francesca De Simone, Lutz Goldmann, Jong-Seok Lee, and Touradj Ebrahimi
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ABSTRACT

Today, several alternatives for compression of digital pictures and video sequences exist to choose from. Beside internationally recognized standard solutions, open access options like the VP8 image and video compression have recently appeared and are gaining popularity. In this paper, we present the methodology and the results of the rate-distortion performance analysis of VP8. The analysis is based on the results of subjective quality assessment experiments, which have been carried out to compare the two algorithms to a set of state of the art image and video compression standards.

Keywords: VP8, WebP, WebM, subjective quality, coding efficiency, rate-distortion curves

1. INTRODUCTION

Multimedia users produce and consume digital images and video sequences in their everyday life and the already large quantity of multimedia material distributed over diverse networks is going to increase in the coming years. Particularly, trends clearly indicate that the video consumption over the Web is on the rise.¹ At the same time, users' demand for increased resolution and higher quality is growing.

In order to deal with this challenging scenario where the network resources will be limited but the amount of data and users will increase, research on image and video compression is continuously developing. Examples of recent efforts to define new international compression standards, with higher coding efficiency than the state of the art solutions, are the JPEG XR image compression standard, approved and published as ITU-T recommendation and ISO/IEC standard in 2010,² and the on-going activities of 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) to define the next generation video coding standard, called High Efficiency Video Coding (HEVC).³

Additionally, while during the last years proprietary technologies such as Apple QuickTime, Microsoft Silverlight and Adobe Flash, which allow video visualizations on the Web, have become popular, the World Wide Web Consortium is currently developing a standard, HTML 5,⁴ that provides the enhanced functionality to embed non proprietary video formats in a web page. This allows users to view video streams embedded in a web page without a specific video player, simplifying the access to video resources.

Considering the need for efficient image and video codecs to optimize resource consumption while ensuring high quality data, as well as the need for a simplified access to multimedia content on the Web, Google recently proposed an open, royalty-free, image and video file format, called WebP for the images and WebM for the audio visual sequences.⁵ This format has been specifically designed for the Web and is quickly gaining popularity, being natively supported in an increasing number of Web browsers. The visual data contained in a WebP or WebM file is encoded using the VP8 open source video compression algorithm.⁶

Only a few existing studies have evaluated the coding efficiency performance of VP8 with respect to other emerging image and video codecs. Image and video coding benchmarks are important for assessing the operational rate distortion performance of codecs and allow for informed decisions on technology deployment and

Further author information: (Send correspondence to Francesca De Simone)

Francesca De Simone: E-mail: francesca.desimone@epfl.ch

Lutz Goldmann: E-mail: lutz.goldmann@epfl.ch

Jong-Seok Lee: E-mail: jong-seok.lee@epfl.ch

Touradj Ebrahimi: E-mail: touradj.ebrahimi@epfl.ch

bandwidth provisioning in multimedia oriented networks and systems. They are conducted either during the development and standardization of new codecs to assess the improved coding efficiency with respect to other existing solutions^{7,8} or after the adoption to compare different codec implementations with each other.

The latest instance of the yearly video coding benchmark conducted by the Moscow State University (MSU)⁹ and their previous study¹⁰ compared different H.264/AVC implementations and VP8, using both compressed and uncompressed source video sequences. At the best of our knowledge, the only study directly comparing both VP8 and HEVC with H.264/AVC is that conducted by Ohworiole and Andreopoulos,¹¹ which considers uncompressed source video sequences and PSNR and SSIM for performance evaluation. Considering the image compression, an extensive study of the performance of WebP in comparison to JPEG has been performed by Google,¹² using SSIM and uncompressed source images, as well as PSNR and JPEG compressed source images.

In all these works the visual quality of the coded material has been evaluated through objective metrics. While these metrics are a valuable tool for automatic codec optimization and evaluation, their correlation with the perceived quality by a human observer is often limited. Therefore, it is also essential to compare emerging and well established codecs through formal subjective quality tests.

In this paper, we present a rate-distortion performance analysis of VP8 based on the results of subjective quality assessment experiments. The VP8 image compression algorithm has been compared to a set of state of the art image compression standards, namely JPEG, JPEG 2000, and JPEG XR. The VP8 video compression algorithm has been compared to H.264/AVC MPEG-4 and to the current version of the HEVC algorithm, as implemented in the latest release of its reference software. To the best of our knowledge this is the first joint assessment of emerging image and video codecs that compares the subjective quality of VP8 to well established (JPEG, JPEG 2000, H.264/AVC) and new coding technologies (JPEG XR, HEVC). The study focuses on the 4:2:0 compression of still and moving pictures, considering uncompressed 4:4:4 source photographic images at high resolution and uncompressed 4:4:4 source natural video sequences at a typical web resolution.

The coding algorithms compared in our study and the main configuration parameters used to produce the test materials are briefly described in section 2*. The test environment, the selected dataset of images and video sequences and the adopted test methodologies are presented in section 3. Finally, section 4 discusses the results of the experiment, while conclusive remarks are drawn in section 5.

2. CODECS AND CONFIGURATIONS

2.1 Image codecs

2.1.1 WebP

WebP is a recent image format developed and sponsored by Google. A WebP file contains an image coded by block-based predictive coding. This image coding strategy is used to code the key frames of a video sequence in the *VP8* codec,⁶ an open and claimed royalty free video codec initially developed by On2 Technologies and later purchased and released by Google. For this work the command line tools *cwebp* and *dwebp*,⁵ version 0.1.2, were used to encode and decode the images. Two coding configurations were used, varying the coding *quality factor* to reach the target bit rates detailed in section 3.2: the default configuration and a second configuration using the provided *photo* preset (we will refer to this configuration as *webp(ps)* in the rest of the paper).

2.1.2 JPEG

*JPEG*¹³ is a block-based image compression standard developed in 1992. *JPEG* is still the most common image compression algorithm used by digital cameras and for storing and transmitting images on the web. The JPEG compressed images used in this study has been produced using the *IJG* implementation[†], version 8c. The images have been coded in baseline profile and the target coding bit rate has been controlled by varying the *quality factor* input parameter.

*When not specified, default parameters have been used. Please contact the authors if you are interested in the detailed command lines and configuration files used for producing the test material.

[†]<http://ijg.org/>

2.1.3 JPEG 2000

*JPEG 2000*¹⁴ is a wavelet-based compression standard for still images, sometimes also used also for frame-based compression of image sequences, such as those in digital cinema. Developed after the *JPEG* standard, *JPEG 2000* significantly outperforms *JPEG* in terms of compression efficiency and offers a large number of features useful in multimedia applications. In this study, the *Kakadu* implementation[‡], version 6.4.1, was used. In order to perform 4:2:0 encoding, the RGB images have been pre-processed, applying RGB to YCbCr color space conversion and then downsampling the chrominance components. The inverse procedure has been applied to the decoded image components to obtain the final 4:4:4 RGB decoded image. The rate control option has been used to encode the images at the target coding bit rates detailed in section 3.2.

2.1.4 JPEG XR

*JPEG XR*² is the latest international standard for image compression, approved and published by ITU-T and ISO/IEC in 2009. Based on the *HD Photo* compression algorithm developed by Microsoft, *JPEG XR* block-based compression uses many of the same fundamental building blocks as in other traditional image and video compression schemes (e.g. color conversion, block-based transform, quantization, coefficient scanning and entropy coding). As major differences with respect to *JPEG*, a reversible Lapped Bi-orthogonal Transform (LBT) and an alternative coefficient coding approach are used. Existing studies on the performance evaluation of *JPEG XR* show that the new standard achieves significantly better compression efficiency than *JPEG*, with overall performance slightly below or comparable to *JPEG 2000*.⁷ Two implementations of the *JPEG XR* codec have been considered. The first implementation is the *JPEG XR* reference software[§], version 1.20. The second one is that provided by Microsoft for the study in⁷ and includes a pre-processing tool to define content adaptive quantization tables¹⁵ (we will refer to this configuration as *xr(ms)* in the rest of the paper). For the *JPEG XR* reference software implementation, the quantization steps for the chrominance channels were derived from the luminance channel quantization steps. For both implementations 4:2:0 encoding has been set. Similarly to *JPEG*, the target coding bit rate has been controlled by varying the *quality factor* input parameter.

2.2 Video codecs

2.2.1 WebM

WebM is an audio-visual format recently developed and sponsored by Google. A *WebM* file consists of a VP8 coded video stream and a Vorbis coded audio stream multiplexed into a Matroska container. The *VP8* video codec⁶ includes similar coding tools than *H.264/AVC* and some alternative tools such as adaptive mixing strategies for artificial reference frames, processor adaptive real time encoding and a low complexity loop filter. For this work the *libvpx* implementation,⁵ version 0.9.6, which offers encoding and decoding functionality, was considered. For optimal rate-distortion (RD) performance, the constant quality bit rate (CBR) mode configuration best performing in¹⁰ was used.

2.2.2 H.264/AVC

The *H.264/AVC* video coding standard¹⁶ was completed by the Joint Video Team (JVT) in 2003 and formally standardized as ITU-T H.264 and ISO/IEC MPEG-4 Part 10. It is currently one of the most commonly used video codecs for recording, compression and distribution of high definition video in a large variety of applications including TV broadcast, video conferencing, web video, and Blu-Ray. With respect to previous video coding standards, such as *MPEG-2* and *MPEG-4 Part 2*, a bit rate gain of more than 50% is achieved through a set of advanced coding tools such as multi-frame variable block size motion compensated prediction, advanced context-based entropy coding, advanced temporal prediction structures and adaptive in-loop deblocking filter. Two implementations have been considered in this study: the open source *x264* implementation[¶], revision r2019, and the *JM* reference software for the H.264/AVC standard^{||}, version 18.0.

[‡]<http://www.kakadusoftware.com/>

[§]<http://www.iso.org>

[¶]<http://www.videolan.org/developers/x264.html>

^{||}<http://iphome.hhi.de/suehring/tml/>

x264 is the most mature open source implementation of the *H.264/AVC* standard and comparable to the best commercial implementations with respect to RD performance. Since it contains only an encoder which can be configured through CLI, the JM decoder was used. The constant bit rate (CBR) mode configuration best performing in¹⁰ was used.

In order to compare *H.264/AVC* and *HEVC*, a non-optimized implementation of the H.264/AVC standard, more comparable to the current implementation of HEVC, the *JM* software was also used in our study. Particularly, a coding configuration satisfying the random access scenario, which could be replicated in a similar configuration of *HEVC*, was selected. This configuration is called alpha anchor configuration.¹⁷ Quantization parameter (QP) based rate control has been considered.

2.2.3 HEVC

Due to the increasing demand for more efficient and flexible video coding solutions beyond *H.264/AVC*, the Joint Collaborative Team on Video Coding (JCT-VC) has recently started to develop a new video coding standard known as high efficiency video coding (*HEVC*). The first evaluations have shown that bit rate gains up to 50% can be achieved with respect to *H.264/AVC*. This gain is achieved thanks to extended or new coding tools, such as larger block sizes with flexible subpartitioning, intra picture prediction from adjacent prediction units, motion vector competition and hierarchical variable length coding. For this work *HM*, the HEVC test model **, version 3.2, was considered. The high efficiency configuration satisfying the random access scenario¹⁷ and quantization parameter (QP) based rate control were used.

3. SUBJECTIVE QUALITY EVALUATION

3.1 Environment and equipment

The subjective tests were performed at the MMSPG test environment,⁸ equipped with three high quality LCD monitors (Eizo CG301W) with native resolution of 2560x1600 pixels, gray-to-gray response time of 6 ms and black-white-black response time of 12 ms. The monitors were calibrated using an EyeOne Display2 color calibration device according to the following profile: sRGB Gamut, D65 white point, 120 cd/m² brightness and minimum black level. The room was equipped with a controlled lighting system that consisted of neon lamps with 6500 K color temperature, while the color of all the background walls and curtains present in the test area was mid grey. The illumination level measured on the screens was 30 lux and the ambient black level was 0.5 cd/m². It is worth mentioning that this same environment was used for evaluation and selection of best performing proposals submitted for HEVC standardization.⁸

3.2 Datasets

3.2.1 Image dataset

For the image codec comparison, 8 images from the JPEG XR evaluation Dataset ^{††} have been used. They are shown in figure 1. All the images have a resolution of 1280x1600 pixels and are available in 4:4:4 RGB uncompressed format. Two images have been used to perform the training of the subjects, while the remaining six have been used as test material. This set of images was coded using the 4 codecs and 6 different coding configurations described in section 2.1. The following 5 coding bit per pixel (bpp) values were selected as target bpp values to be analyzed: 0.125, 0.25, 0.5, 0.75, and 1 bpp. For content *bike*, *JPEG* and *VP8* (in both configurations) did not reach the target 0.125 bpp value. For content *p01*, *VP8*(in both configurations) did not reach the target 0.125 bpp value. Thus, this resulted in a final test set of 175 coded images, which have been used for the subjective evaluation.

**https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

††<http://mmspg.epfl.ch/iqa>

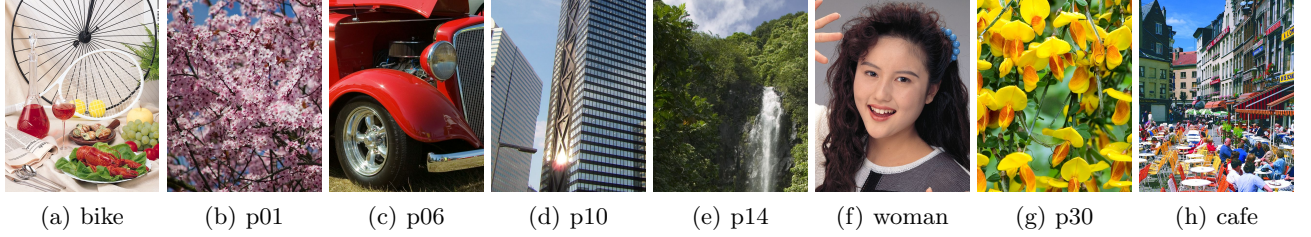


Figure 1. Image dataset: test contents (a-f) and training contents (g-h).



Figure 2. Video dataset: test contents (a-d) and training content (e).

3.2.2 Video dataset

For the video codec comparison, 5 videos from the VQEG HDTV SVT Dataset ^{††} have been used. A representative frame of each sequence is shown in figure 2. The original videos are available in various resolutions, scans and frame rates (2160p50, 1080p50, 1080i25, 720p50, 576i25) and have a duration of 10 seconds. The 1080p50 version of the videos was used and spatially and temporally subsampled to a typical web resolution of 854x480 pixels and a frame rate of 25 fps. One sequence was used for training of the subjects, while the remaining four have been used as test material. This set of sequences was coded using the 4 codecs described in section 2.2. The following 5 coding bit rates were selected as target bit rates of the study: 250, 500, 750, 1250, and 2250 kbps. For content *CrowdRun* and *DucksTakeOff*, VP8 (in both configurations) did not reach the two lowest bit rates (250 and 500 kbps), thus the closes bit rate reached by the codec was selected instead. This resulted in a final test set of 80 coded video sequences, which have been used for the subjective evaluation.

3.3 Test methodology

3.3.1 Double Stimulus Continuous Quality Scale image evaluation

The subjective quality evaluation to compare the image compression algorithms described in section 2.1 has been performed following the methodology proposed in ⁷. As an adaptation of the double-stimulus continuous quality scale (DSCQS) method for video quality evaluation, ¹⁸ the selected method implies that two images are displayed simultaneously by splitting the screen horizontally into two parts. One of the two images is always the reference, unimpaired, image. The other is the test image, which in this study is a compressed version of the reference. The subject is not told about the presence of the reference in each pair and, after the visualization, is asked to rate the quality of both stimuli, using for each a continuous quality scale ranging from 0 to 100, associated with 5 distinct quality levels (*Bad*, *Poor*, *Fair*, *Good*, *Excellent*). The position of the reference image is randomly selected at each visualization.

Since each test session was run with a group of 2 subjects, each sitting in front of one display receiving input from the same video server, the image visualization time has been fixed to 17 seconds and the rating were collected using paper scoring sheet. After 12 seconds of visualization, a message appearing at the bottom of each displayed image asked the subjects to enter their rates. From the instant the message appeared, the subjects had 5 seconds to enter their rates before the visualization of the next pair of images started. These visualization and voting times allowed a detailed exploration of the high resolution content while assuring a reasonable duration of the test. A picture of the graphical user interface during the rating time is shown in figure 3 (a). The selected rating scale is shown in figure 3(b).

^{††}ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/

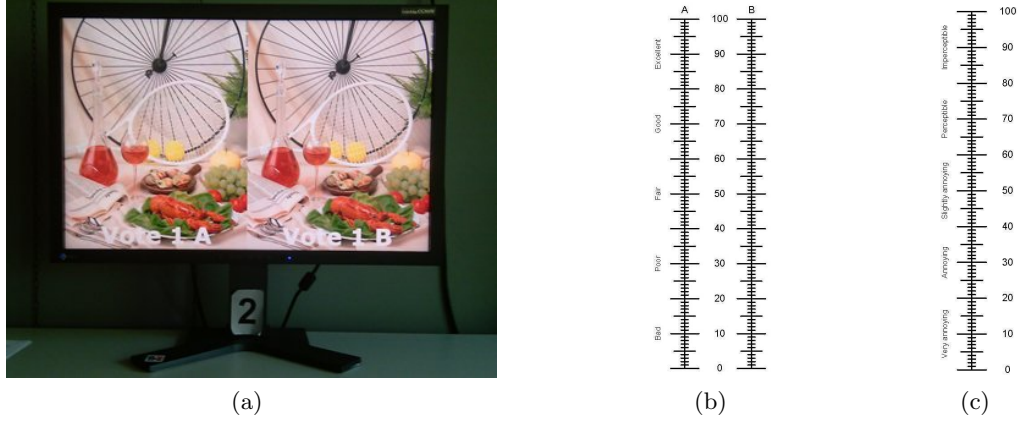


Figure 3. Subjective evaluation: picture of the graphical user interface used for the DSCQS image quality evaluation, during the rating time (a); rating scales used for the DSCQS method (b) and the DSIS method (c).

3.3.2 Double Stimulus Impairment Scale video evaluation

For the video quality evaluation, the standard Double Stimulus Impairment Scale (DSIS) methodology¹⁸ has been selected. According to this method, pairs of sequences, i.e. stimuli A and B, are sequentially presented to the subject and she/he is asked to rate the quality of the second stimulus. The subject is told about the presence of the reference video, having the best expected quality, as stimulus A and she/he is asked to rate the level of annoyance of the visual defects that she/he observes in stimulus B (*Very annoying*, *Annoying*, *Slightly annoying*, *Perceptible*, *Imperceptible*).

As for the image quality evaluation, each test session was run with a group of 2 subjects, each sitting in front of one display receiving input from the same high performance video server, and the ratings were collected using paper scoring sheet. After the visualization of each pair of video sequences, a 5 seconds long grey screen appeared, with a message asking the subjects to enter their rates. The used rating scale is shown in figure 3(c).

3.3.3 Training, multiple sessions and subjects

For both the image and video quality subjective evaluations, before the test, oral instructions were provided to the subject to explain his/her task. Additionally, a training session was performed to allow the viewer to familiarize with the assessment procedure. As detailed in 3.2, the contents shown during the training session were not used in the test. The training samples have been manually selected by an expert viewer so that the quality of each sample was representative of one categorical quality level on the rating scale.

Since the total number of test samples, both for the image and the video evaluations, was too large for a single test session, the image quality test was split into 4 sessions of approximately 14 minutes each, the video quality test in 3 sessions of approximately 13 minutes each. After each session each subject took a 10 minutes break, before starting the next session. Each session included test material corresponding to all the different contents, all the codecs under analysis, and only a subset of the bit rates, which were uniformly distributed across the sessions.

Three dummy pairs were included at the beginning of the first and the third image sessions and at the beginning of the first video session, in order to stabilize the viewer's judgment, and one stimuli pair was repeated within each test session, in order to check the reliability of subject's rating. Finally, for each session, a different permutation of the same stimuli list was used for each group of 3 to 4 subjects.

A total of 18 people, 5 women and 11 men, with an average age of 24 years old, took part to the image quality test, completing all the test sessions. A total of 18 people, 9 women and 9 men, with an average age of 24 years old, took part to the video quality test, completing all the test sessions. All the participants were naive subjects and reported normal or corrected to normal visual acuity and color vision.

3.4 Score processing and analysis

The scores resulting from the image quality test were processed separately from those resulting from the video quality test. The collected subjective scores were processed according to the procedure described in.⁸ The results of different groups of subjects were merged before performing the statistical analysis of the data.

The outlier detection algorithm described in⁸ was applied to the scores of each session, in order to detect and remove subjects whose scores appear to deviate strongly from the other scores in the same session. A subject was considered as an outlier, and thus all his/her scores were removed from the results of the session, if more than 20% of his/her scores over the session were outliers. None of the subjects were detected as outlier for any of the image test sessions, neither for the video test sessions.

After the outlier detection, statistical measures were computed to describe the score distribution across the subjects for each of the test conditions (combination of content, coding condition and bit rate), as described in.⁸ For the DSIS methodology, the mean opinion score (MOS) was computed for each test condition. For the DSCQS methodology, the differential mean opinion score (DMOS) was computed for each test condition. The DMOS values, in the range $[-100, 0]$, were converted to MOS values in the range $[0, 100]$, to uniform the range of the rate-distortion curves of the image and video codec comparison and facilitate the interpretation of the results. The 95% confidence intervals (CI) for the MOS values were computed using the Student's t-distribution.

Finally, a multiple comparison analysis¹⁹ was performed, in order to identify the statistically significant differences among the MOS values obtained for different codecs and the same bit rate condition.

4. RESULTS AND DISCUSSION

4.1 Image codec comparison results

Figure 4 shows, for each test content, the rate-distortion plots with the MOS and CI values obtained after processing the subjective results. From the rate distortion curves it is possible to have an overall impression of the performance of the different codecs. In general WebP shows comparable performance to JPEG 2000 and JPEG XR. The only exception is for image *p01*, where for low bit rate the quality of the WebP images has been judged as lower than the others, apart from JPEG which is always outperformed by the other algorithms for bit rate values lower than 0.5 bpp.

The preliminary codec performance comparison based on the analysis of the RD curves is confirmed by the results of the multiple comparison analysis. Each checkboard plot in figure 5 shows the number of rejections of the null hypothesis that the MOS values of two codecs are the same for the same bpp value. When the rejection number is equal to zero it means that, for the bpp under analysis, the two codecs always have the same performance. As the opposite case, when the rejection rate is maximum, i.e. equal to 6, it means that the two codecs never have the same performance. It should be noticed that for the 0.125 bpp case, the JPEG sample for content *bike* was not available and the VP8 samples were not available for both content *bike* and *p01*, since these codecs did not reach the target bpp for these contents. Thus the results shown in figure 5(a) for JPEG and both configurations of WebP, are computed over a set of only 5 and 4 pictures, respectively. For 0.125 bpp value, JPEG 2000 shows overall the best performance, outperforming particularly JPEG XR Microsoft implementation for 5 out of 6 images, as well as JPEG for 3 out of 5 pictures. For 0.250 bpp value, apart from JPEG, all the other codecs have comparable performance. For 0.5 bpp value, the JPEG XR reference software implementation is usually outperformed by WebP and JPEG 2000, which have the same performance. Finally for bpp values equal or greater than 0.75 bpp all the codecs have comparable performance.

4.2 Video codec comparison results

As for the image quality comparison, figure 6 shows, for each test content, the RD plots with the MOS and CI values obtained after processing the subjective results of the video quality evaluation. While CrowdRun and DuckTakeOff are difficult contents for all the codecs, the perceived quality saturates quickly for IntoTree and OldTownCross. Overall, from the RD curves, HM and x264 show the best performance. For CrowdRun and DuckTakeOff, apart from WebM, all the codecs show comparable performance. For IntoTree and OldTownCross at low bit rate, JM is usually outperformed by all the other codecs, which show smaller difference. The fact that

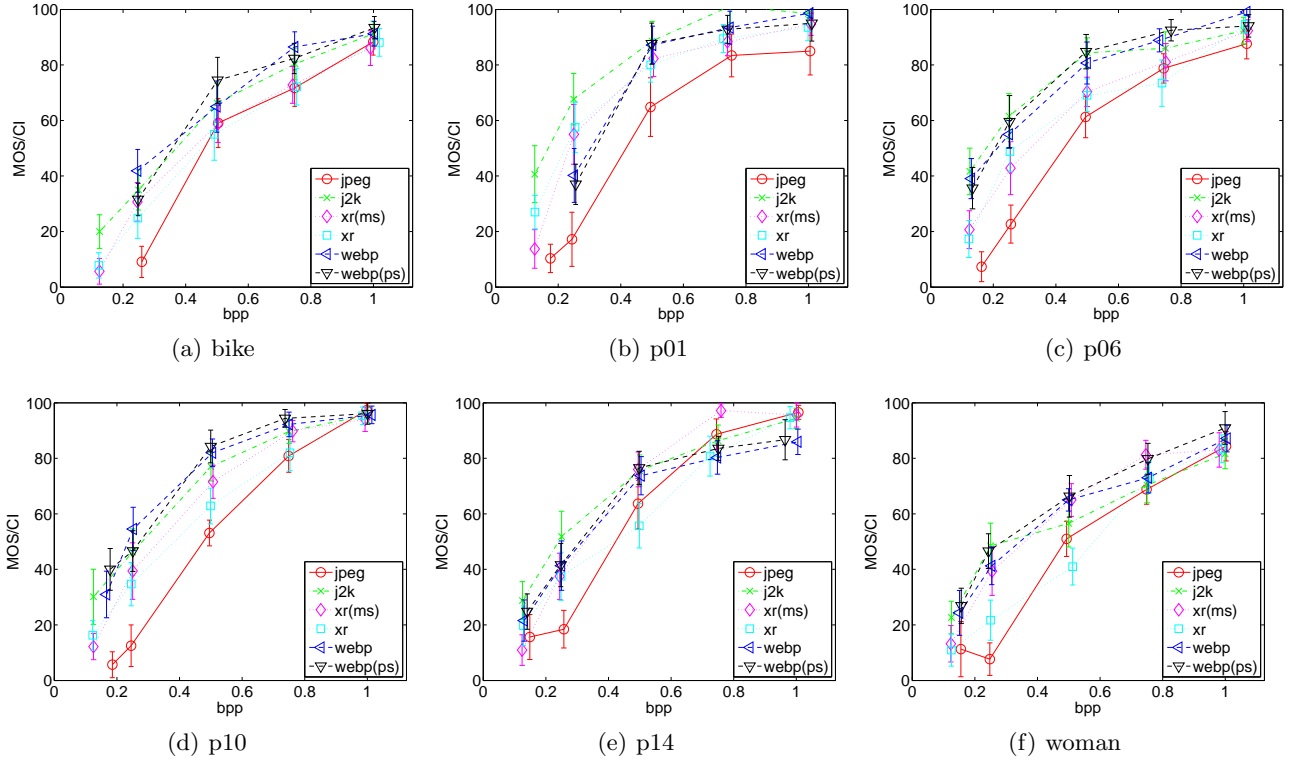


Figure 4. RD curves for the test images.

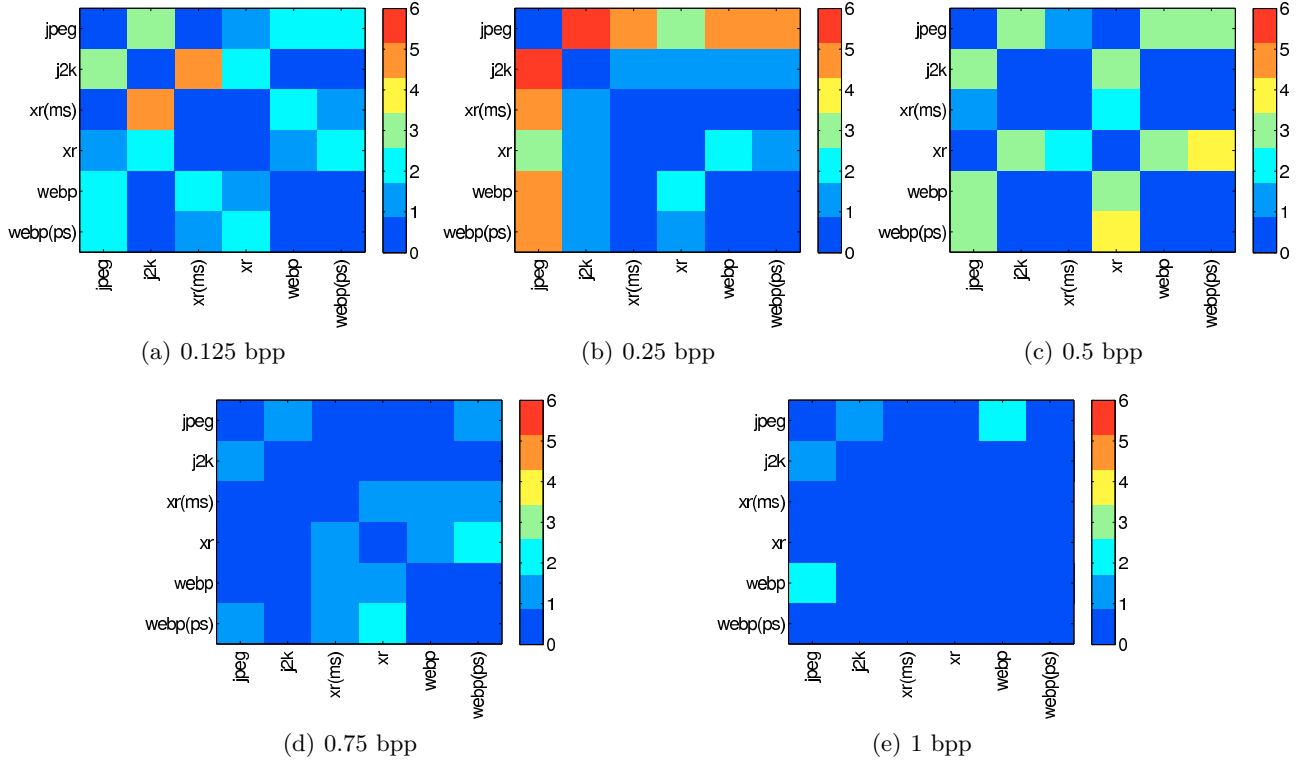


Figure 5. Results of the multiple comparison analysis as number of rejections of the null hypothesis that the MOS values of two codecs are statistically the same for the same bpp value, over the 6 contents. For the 0.125 bpp case (a), the results for JPEG and both configurations of WebP are computed over a set of only 5 and 4 pictures, respectively.

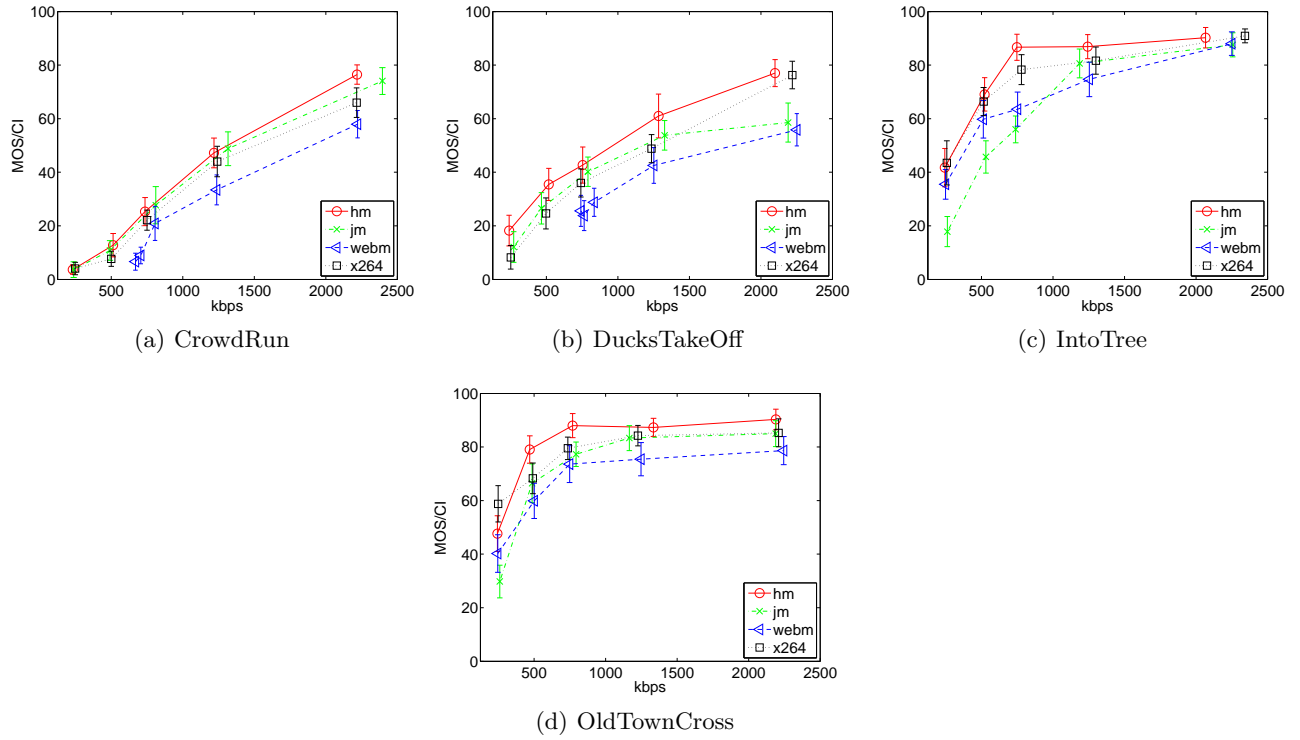


Figure 6. RD curves for the test sequences.

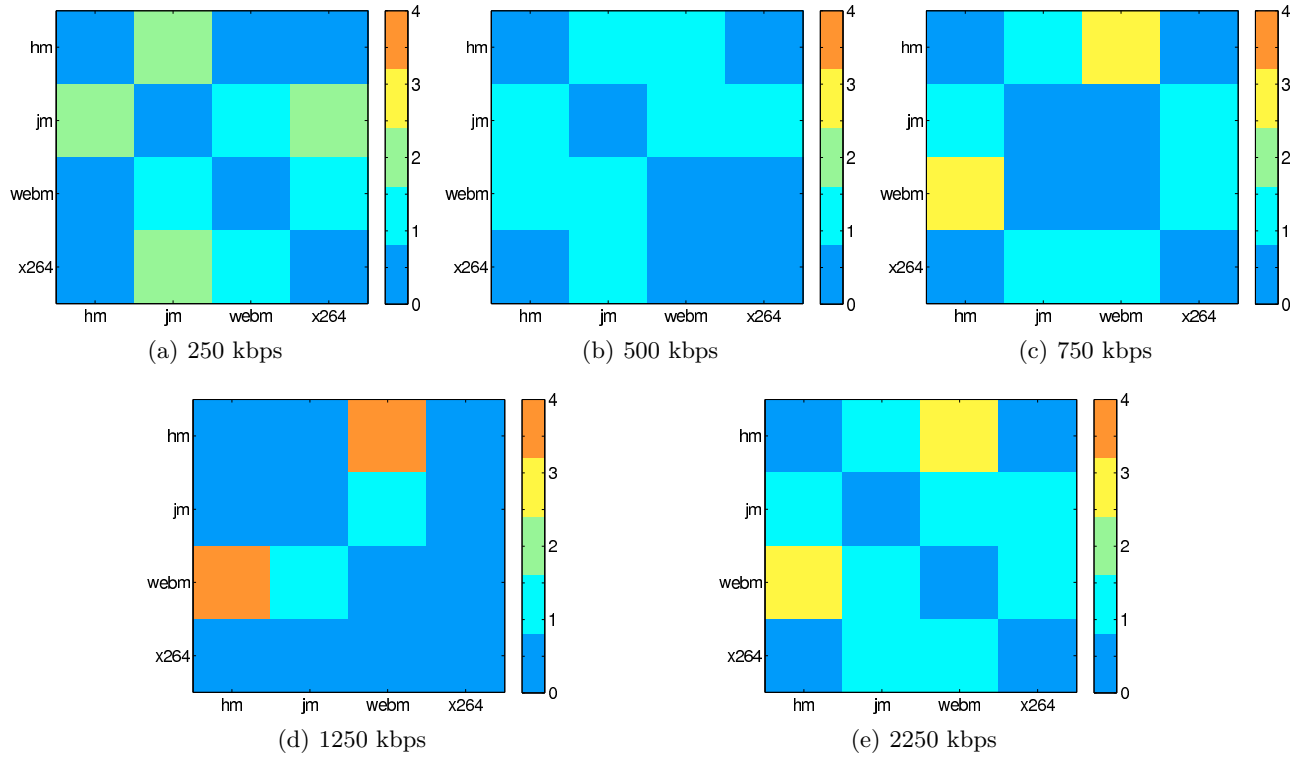


Figure 7. Results of the multiple comparison analysis as number of rejections of the null hypothesis that the MOS values of two codecs are statistically the same for the same bps value, over the 4 contents. For the 250 kbps case (a) and the 500 kbps case (b), the results for WebM are computed over a set of only 2 sequences.

the difference in performance for HM, x264 and WebM becomes smaller for simpler contents and low bit rate conditions may be expected considering that both HM and JM coded the sequence in random access mode, thus, will probably use a large portion of the bit rate for coding the I frames. x264 and WebM instead are not used in random access configuration, therefore can better allocate the bits over all the frames of the video sequence.

The results of the multiple comparison analysis are shown in figure 7 for each bit rate value separately, over the entire set of 4 sequences under analysis. It should be noticed that for the 250 kbps case and the 500 kbps case, the VP8 samples for content *CrowdRun* and *DucksTakeOff* were not available, since the codec did not reach the target bit rate for these contents. So, the results shown in figure 7 (a) and (b) for VP8 are computed over a set of only 2 pictures. respectively. As already noticeable in the RD plots, WebM is usually outperformed by HM, especially for bit rates greater than 500 kbps, but has comparable performance to x264 and JM.

5. CONCLUSION

In this paper a rate-distortion performance analysis of VP8 image and video compression, based on the results of formal subjective quality evaluation, has been described. The VP8 image compression has been compared to three state of the art standards for image compression, namely JPEG, JPEG 2000 and JPEG XR, over a set of 6 different photographic contents. The original images were uncompressed RGB 4:4:4 high resolution pictures and 4:2:0 coding was considered. The VP8 video compression has been compared to two state of the art standards for video compression, namely the well established H.264/AVC and the video coding standard currently under definition, HEVC, over a set of 4 different video sequences. The original sequences were uncompressed RGB 4:4:4 SDTV sequences, and 4:2:0 coding was considered.

Two groups of eighteen naive subjects took part to the image and video quality evaluation experiments, respectively. Each subject participating to the image quality test took part to 4 separated test sessions, scoring a total of 175 test images. Each subject participating to the video quality test took part to 3 separated test sessions, scoring a total of 80 test sequences.

A detailed statistical analysis of the subjective results has been performed. The obtained results allow an accurate comparison of the performance of the different codecs for the test conditions selected for this study. Overall, from the results shown in the paper it can be concluded that the VP8 image compression showed performance comparable to JPEG XR and JPEG 2000, all significantly outperforming JPEG compression. For video compression, the performance of VP8 were competitive with x264, while, interestingly, the new HEVC technology under definition usually showed the best performance. Finally, for some contents, both for image and video compression, it could be noticed that the current implementation of VP8 seemed to have difficulties reaching low bit rates conditions that most other codecs reached.

Future studies will consider other resolutions of the same test material, as well as other subjective evaluation methodologies.

ACKNOWLEDGMENTS

The work presented in this paper was supported by the Swiss National Foundation for Scientific Research in the framework of NCCR Interactive Multimodal Information Management (IM2), the European Network of Excellence PetaMedia (FP7/2007-2011) and the COST IC1003 Network of Excellence QUALINET (2010-2014). Additionally, the authors would like to thank Miss Althea Araya for the help in performing the subjective tests and all the people who took part in the experiments.

REFERENCES

- [1] Begen, A., Akgul, T., and Baugher, M., "Watching video over the web: Part 1: Streaming protocols," *Internet Computing, IEEE* (March-April 2011).
- [2] ISO/IEC, "Information technology - jpeg xr image coding system - part 2: Image coding specification," Tech. Rep. 29199-2:2009, ISO/IEC (2009).
- [3] Sullivan, G. J. and Ohm, J.-R., "Recent developments in standardization of high efficiency video coding (HEVC)," in [*SPIE Applications of Digital Image Processing XXXIII*], **7798** (August 2010).

- [4] Vaughan-Nichols, S., “Will HTML 5 restandardize the web?,” *Computer* **43**, 13–15 (April 2010).
- [5] “<http://www.webmproject.org/>.”
- [6] Bankoski, J., Wilkins, P., and Xu, Y., “Technical overview of vp8, an open source video codec for the web,” (2011).
- [7] De Simone, F., Goldmann, L., Baroncini, V., and Ebrahimi, T., “Subjective evaluation of JPEG XR image compression,” in [*SPIE Applications of Digital Image Processing XXXII*], **7443** (August 2009).
- [8] De Simone, F., Goldmann, L., Lee, J. S., and Ebrahimi, T., “Towards High Efficiency Video Coding: Subjective Evaluation of Potential Coding Technologies,” *Journal of Visual Communication and Image Representation* (2011).
- [9] “http://www.compression.ru/video/codec_comparison/h264_2011/.”
- [10] “http://www.compression.ru/video/codec_comparison/h264_2010/vp8_vs_h264.html.”
- [11] Ohwovoriole, E. and Andreopoulos, Y., “Rate-distortion performance of contemporary video codecs,” (2010).
- [12] “http://code.google.com/speed/webp/docs/webp_study.html/.”
- [13] ITU-T, “Information technology - Digital compression and coding of continuous-tone still images - Requirements and guidelines,” Tech. Rep. T.81, ITU-T (1992).
- [14] ITU-T, “Information technology - JPEG 2000 image coding system: Core coding system,” Tech. Rep. T.800, ITU-T (2002).
- [15] Schonberg, D., Sullivan, G. J., Sun, S., and Zhou, Z., “Perceptual encoding optimization for JPEG XR image coding using spatially adaptive quantization step size control,” in [*SPIE Applications of Digital Image Processing XXXII*], **7443** (August 2009).
- [16] Wiegand, T., Sullivan, G., Bjontegaard, G., and Luthra, A., “Overview of the h.264/avc video coding standard,” *Circuits and Systems for Video Technology, IEEE Transactions on* **13**, 560–576 (July 2003).
- [17] JTC1/SC29/WG11, I., “Joint call for proposals in video compression technology,” Tech. Rep. N11113.
- [18] ITU-R, “Methodology for the subjective assessment of the quality of television pictures,” Tech. Rep. ITU-R BT.500-11, ITU-R (2002).
- [19] Snedecor, G. W. and Cochran, W. G., [*Statistical Methods*], Iowa State University Press (1989).