The 7th International Telecommunications Symposium (ITS 2010)

Quality Performance of LD, SD and HD Video Encoded with Dirac and H.264/AVC

Wyllian B. da Silva, Keiko V. O. Fonseca and Alexandre de A. P. Pohl
LCD/CPGEI, Universidade Tecnológica Federal do Paraná (UTFPR), Curitiba, PR, Brasil
wyllianbs@gmail.com, keiko@utfpr.edu.br, pohl@utfpr.edu.br

Abstract— **This paper compares the video quality performance of low definition (LD), standard definition (SD) and high definition (HD) video signals encoded with H.264/AVC and Dirac using full-reference (PSNR and SSIM) and no-reference metrics (JPEG-NR) for different bit rates. It is shown that H.264 outperforms Dirac concerning quality for all resolutions.**

Keywords — Dirac, H.264/AVC, Digital TV.

I. INTRODUCTION

The increasing number of digital video services and growing popularity of high definition and 3D television are creating an enormous need for higher coding efficiency. Since the development of the early standards, such as MPEG-1, evolving to the enabling MPEG-2 technology that is still widely used for the transmission of low (LD), standard (SD) and high definition (HD) TV signals, there is a constant search for new video encoders. Moreover, the transmission of good quality video over much lower data rate media, such as in wireless and access networks, has enhanced the need for efficient and flexible coding. As such, H.264 (AVC) has emerged in 2003 as the successor of MPEG-2 after a joint development effort by the ITU Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). Its video coding layer (VLC) is based on the conventional block-based motion-compensated hybrid video concept, but with important new features, such as the enhanced motion-prediction capability, the use of a small block-size exact match-transform, the adaptive in-loop deblocking filter and enhanced entropy coding methods [1]. As a promising alternative to H.264, the Dirac concept was developed by the British Broadcast Corp. (BBC) R&D team as an open platform that can be used by anyone without paying license fees [2, 3]. It is also based on the conventional block-based motion-compensated hybrid video concept, but differs from H.264, for example, by the use of the discrete wavelet transforms, DWT, (rather than the discrete cosine transform, DCT).

As users face the challenge of working with such encoders, the quest for the quality of encoded video resulting from these two technologies become crucial, particularly at low bit rates, which are most used for mobile devices. This way, we have employed the H.264 and Dirac encoder for compression of three video signals with different content and with different resolutions (low, standard and high definition) and have used full-reference and no-reference metrics to evaluate the video quality at different bit rates. As full-reference metrics the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity (SSIM) index were employed. And for the no-reference we used a perceptual quality metric, initially developed for still images (JPEG), but now modified to enable the evaluation of video frames. The results presented in this paper attempts to complement two other works [4, 5] in terms of adding information concerning the codification and evaluation of low-definition video signals and the employment of a no-reference technique, whose results are compared to results obtained with PSNR and SSIM.

The paper is organized as follows: Section II describes the main features of the Dirac and H.264 video codecs. Section III discusses the objective video quality evaluation and corresponding results are show in Section IV, followed by the conclusion in Section V.

II. DIRAC AND H.264/AVC VIDEO CODECS

The H.264 encoder explores techniques such as the intra prediction with spatial correlation that reduces redundancy. In the Baseline profile it allows the choice of four 16x16 and nine 4x4 block size modes, respectively. H.264 utilizes the integer DCT to accelerate the transform calculation and makes use of the Integer Cosine Transform (ICT) with left or right shift of the coefficients to replace multiplication and division operations. This way, it facilitates the transform implementation.

For motion compensation H.264 employs several techniques, which use up to seven variable block sizes. The choice of block sizes depends on the level of details in the frames. For a high detailed block, the small mode size is more appropriate. However, for motionless background, the encoder can choose 16x16 block size coding. However, during this process the encoder may introduce blocking artifacts. For mitigating the impact of the resulting artifacts, an effective deblocking filter is used. Moreover, the encoder uses multiple reference frames and their number can vary from 1 up to 16. Entropy coding methods such as Context-based Adaptive Binary Arithmetic Coding (CABAC) and Context-Adaptive Variable Length Coding (CAVLC) are also used in H.264, which are more efficient than the conventional Huffman technique [6, 7].

On the other hand, the Dirac video codec is an open-source for multiple resolutions and interlaced and progressive formats. Dirac uses the wavelet transform with separable wavelet filters, whose coefficients are divided into sub-bands and quantized employing rate-distortion optimization (RDO) quantizers.

Table I show the basic settings used in both codecs for the comparison of the video quality performance within this work.

TABLE I. CONFIGURATION OF THE X264 AND DIRAC CODES.

Parameter	x264	Dirac
Bit rate (kbps)	32 to 8000	384 to 8000
FPS	30	30
Profile/level	LD: baseline 1.0 SD/HD: high 5.1	–
Number B-frames	2	–
Size of GOP	10	–
Separation of L1-frames	–	2
Number of L1-frames	–	2
Progressive mode	LD/SD	LD/SD
Interlaced mode	HD	HD
Adaptive spatial transform size	8x8	–
Chroma sampling format	4:2:0	4:2:0
Entropy coding	CABAC	Arithmetic coding
Quality factor (QF)	default-	default
CPU capabilities	MMX2, SSE2Fast, SSSE3, FastShuffle, SSE4.1, Cache64	not available

The great advantage of DWT as compared to DCT-based H.264 is that it preserves the fine details. However, DWT may cause ringing and blurring artifacts which are reduced with filters of compact impulse response, called Daubechies filters. Moreover, Dirac uses I-frame, L1 and L2 frames in its group of pictures (GOP). Intra frames (I-frames) are used as reference frames by successive frames into GOP structures. L1 and L2 are both inter frames. L1 frames can be used as temporal reference by L2 frames. Dirac adopts a four level wavelet transform. The filter produces four sub-bands called Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH). Each sub-band contains coefficients that represent specific areas within the frame. Dirac also employs arithmetic coding in its entropy coding stage to produce efficient variable length codes.

III. OBJECTIVE VIDEO QUALITY EVALUATION

This work uses two methods for quality assessment. The full-reference method requires the reference frame (original) and the encoded frame for the pixel to pixel comparison. Two metrics were employed: the Mean Square Error (MSE)/PSNR and SSIM.

The PSNR (dB) of a frame is calculated as

$$PSNR_f = 20\log\left(\frac{v_{peak}}{MSE}\right) \tag{1}$$

where ν_{peak} is $2^k - 1$ and k is the number of bits per pixel. The MSE values vary from 0 to 255. To an increase of MSE corresponds a decrease in PSNR. Typical values for good perceptual quality reported in the literature range between 20 dB and 40 dB [8].

The SSIM compares the local patterns of pixel intensities that have been normalized for luminance and contrast [9, 10]. The quality index of SSIM varies from 0 to 1, where 1 indicates that the encoded frame is the same as the original frame. The SSIM index for a frame with resolution (x,y) is derived from the combination of three parameters: the luminance comparison l(x,y), the contrast comparison c(x,y) and the structure comparison s(x,y). The SSIM is a metric based on Human Visual System (HVS) and it is highly adapted for extracting structural information from a scene of a video content. The SSIM index is expressed as

$$SSIM_f = [l(x,y)]^{\mu}[c(x,y)]^{\sigma}[s(x,y)]^{\omega} \tag{2}$$

where, typically, $\mu = \sigma = \omega = 1$.

The no-reference method uses only the encoded frame for the evaluation. We used the perceptual quality assessment based on JPEG compressed images [11, 12], which we called JPEG-NR index (JPEG No-Reference index). The metric also defines the quality range from 0 to 1. It explores the horizontal and vertical features of the encoded frame to detect blocking and blurring artifacts. The detection of blocking and blurring is obtained by processing the signal from the space-time domain to the frequency domain. The JPEG-NR index is calculated as follows

$$JPEG_f^{NR} = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3} \tag{3}$$

where A is the parameter related to the evaluation of the blurring effect and B is the parameter related to the evaluation of the blocking effect and Z is the mean zero-crossing rate (activity measure) between the horizontal and vertical features. The model parameters α, β, γ_1, γ_2, and γ_3 are obtained from training the video data [11]. The model developed for still images can be extended to evaluate video signals by averaging the index obtained for each frame over the number of frames.

The three video sources used in the analysis are "Riverbed", "Bee and Sunflower" and "Tractor". They are available in high definition (HD) and were obtained from [13]. The first contains 250 frames with little movement, the second contains 500 frames with slow motion of objects ("Bee and Sunflower") and the third shows a greater movement ("Tractor" with 690 frames). By using Matlab tools [14] the HD (1920x1080) format was resized to SD (720x480) and to LD (320x240) format before codification. By the employment of the PSNR technique we noticed that no significant distortion in this conversion step was performed. Figure 1 shows a frame of each of one the videos used in the analysis.

a. b. c.

Figure 1. Frame of: a. Riverbed, b. Sunflower and c. Tractor.

The Intel Core i7 64 bits processor was used for data processing and analysis. The baseline profile is chosen in H.264 for processing the LD (240@30 frames/s) signal and the high profile is used for SD (480@30 frames/s) and for HD (1080@30 frames/s) resolutions. We noticed that the processing time using the x264 code is much faster than the one using the Dirac code. The corresponding codes for the Dirac and H.264 codecs were obtained from [2] and [15], respectively.

Figure 2 shows a typical evolution of PSNR over the number of frames for two different bit rates (1024 kbps and 8000 kbps) of the "tractor" video at LD resolution. At the 1024 kbps rate H.264 presents PSNR values between 34 dB and 40 dB, which are higher than the ones obtained with Dirac, which lay between 30 dB and 35 dB. At the highest bit rate (8000 kbps) H.264 continues to outperform Dirac, with PSNR values between 48 dB and 50 dB. However, for an insight of the performance at increasing bit rates over the video total number of frames, one is required to average the index values obtained frame by frame as seen in Figure 2.

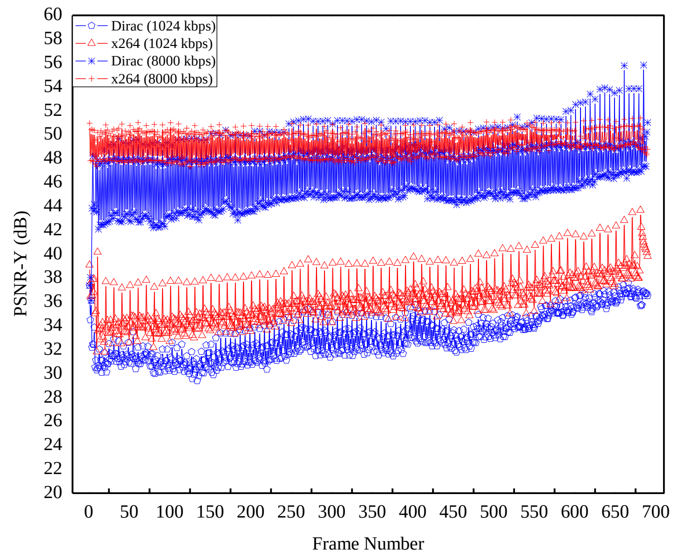

Figure 2. Trend of PSNR (dB) frame by frame with increasing bit rate for the Tractor content at LD resolution.

This way, we have calculated the average of PSNR, SSIM and JPEG-NR index according to

$$\frac{1}{N}\sum_{k=1}^{N} M_k^{frame} \tag{4}$$

where M^{frame} means the metric index (PSNR, SSIM or JPEG-NR) of frame *k*. The sum over N represents the average index obtained over the total number of frames of the video signal.

IV. Results

Figure 3 shows the PSNR performance over the bit rate. For the LD resolution PSNR results show that H.264 outperforms Dirac. A similar behavior is observed for the SD and the HD resolutions, as well. For bit rates over 500 Kbps the PSNR is higher than 27 dB for all resolutions, an indicative that the perceptual image quality is high. Additionally, at increasing bit rates the performance for the LD resolution is higher than for the SD, which is also higher than HD. This confirms the fact that signals of lower resolution at higher bit rates suffers less the impact of the compression process. During the work with Dirac one must be particular careful as the bit rate is reduced [16]. When the encoder is forced to lower rates it delivers a constant quality index, as if Dirac tries to establish a minimum rate that preserves quality. This problem is enhanced as the video resolution increases (some of the curves in Figures 3 and 4 intentionally shows this behavior). Particularly, rates must be set higher than 400 kbps for LD, higher than 1000 kbps for SD and higher than 4000 kbps for HD resolution in order to avoid the problem.

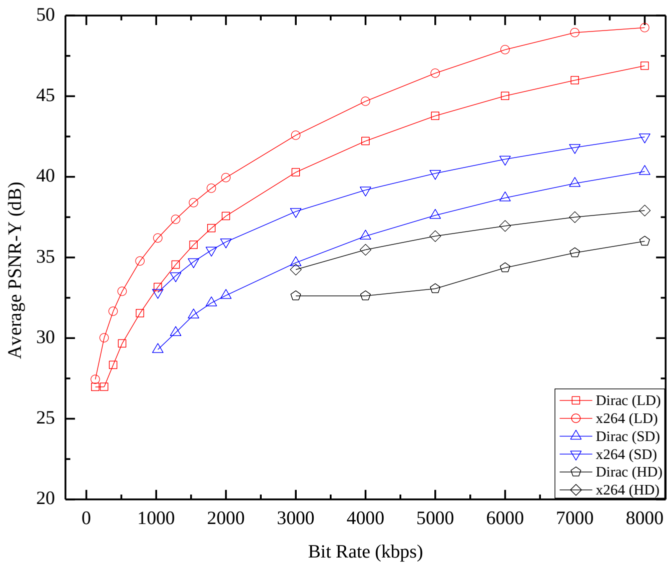

Figure 3. Comparison of average PSNR (dB) of the "Tractor" content after compression at different resolutions (LD, SD and HD).

Figure 4 shows the same comparison using the SSIM index. This index represents the human perceptual quality in a better way since it combines the perception of luminance, contrast and structure of the video. For bit rates over 2000 kbps the index points out to a very high quality (> 0.9) for all resolutions. Again, H.264 outperforms Dirac in all cases. However, for lower bit rates the quality differences can be better distinguished. These rates are particular important for LD transmission in mobile devices. At 500 Kbps, for example, the SSIM index for Dirac is around 0.8 and for H.264 it is 0.9 (LD resolution), a significant difference in terms of the SSIM metric. However, images are still visualized by the HVS without much annoyance.

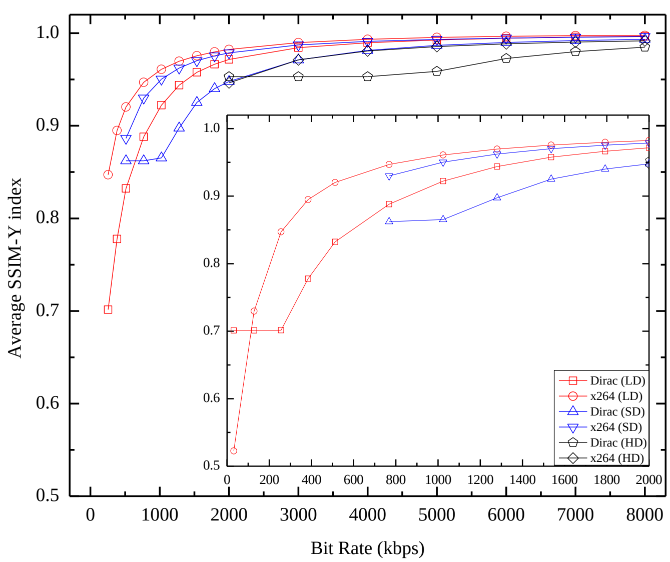

Figure 4. Average SSIM index of video Tractor with resolutions LD, SD and HD.

Figure 5 shows results of the no-reference metric with the JPEG-NR index for the Tractor content, representing resolutions LD, SD and HD.

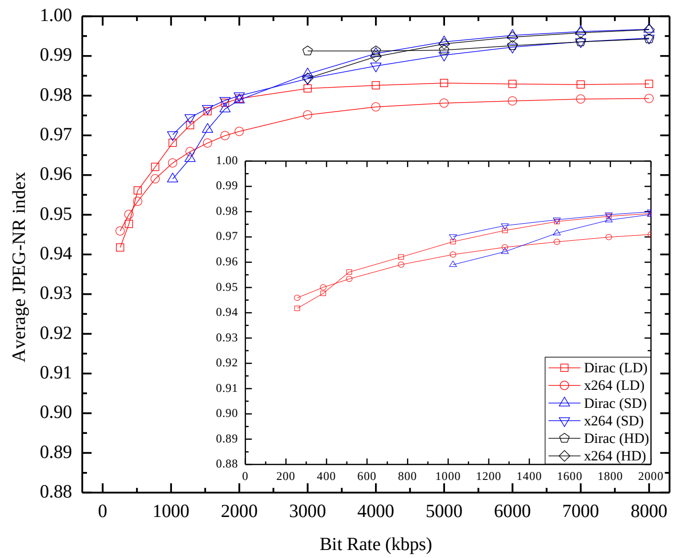

Figure 5. Average JPEG No-Reference of video: Tractor with resolutions LD, SD and HD.

Figure 6 shows the performance comparison with the SSIM index metric for the videos Riverbed, Sunflower and Tractor at the LD resolution.

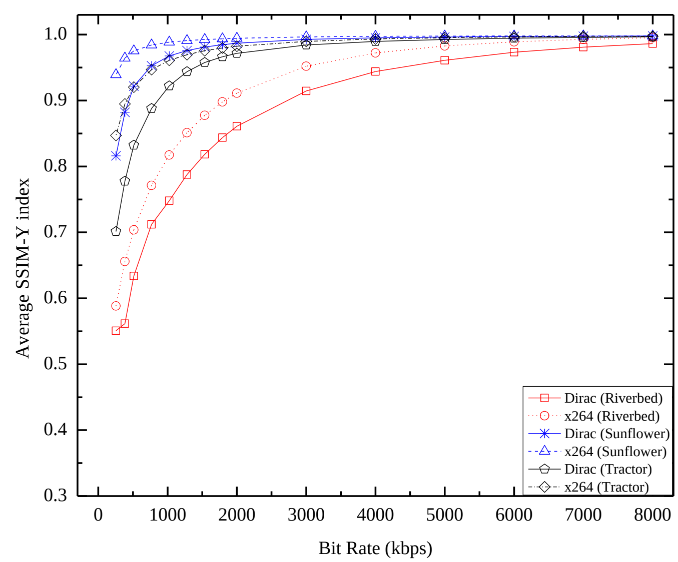

Figure 6. Average SSIM index for three videos with resolutions LD, SD and HD.

This metric shows a different trend than the other two metrics, where Dirac outperforms H.264. This is probably due to the limited training of the used dataset, from which the model parameters α, β, γ_1, γ_2, and γ_3 are obtained. For instance, from 500 kbps to 8 Mbps for the LD resolution, H.264 presents an average JPEG-NR index lower than Dirac. For the SD resolution the x.264 code shows an JPEG-

NR index higher than Dirac up to 2000 kbps, then an inversion occurs and Dirac begins to deliver higher values of JPEG-NR.

For all contents H.264 shows higher SSIM indices than Dirac. Above 3000 kbps the SSIM index is in practice the same for all contents, indicating that no significant changes are perceived by the HVS.

V. Conclusion

This work compared the performance of the H.264 and Dirac in terms of the video quality, which is assessed with full-reference (PSNR and SSIM) and no-reference (JPEG-NR) metrics. Three video signals with different contents and different resolutions (LD, SD and HD) were employed in the analysis. Results indicate that H.264 outperforms Dirac in all cases over the employed bit rate. Other results in the literature confirm this behavior [4]. However, the JPEG no-reference metric used in this work must still be improved in order to deliver results, which better correlate with values given by the PSNR and SSIM metrics. It was also noticed that the x264 code presents a higher speed in the encoding processing because it uses enhanced CPU capabilities (shown in Table I), a feature that the Dirac code still misses.

Acknowledgment

Wyllian Bezerra da Silva thanks CAPES for a PhD scholarship. Authors thank Prof. K.R. Rao of University of Texas at Arlington, USA, for helpful suggestions. This work has been supported by the project “Formação de Pessoal Qualificado em Sistemas de Transmissão de TV Digital no Paraná – Processo 23038.23556/2008-16 AUX-PE-RH-TVD 249/2008” supported by CAPES.

References

[1] T. Wiegand, G, J. Sullivan, *Overview of the H.264/AVC Video Coding Standard*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.

[2] BBC Research and Development. Dirac video codec. <http:dirac.sourceforge.net>.

[3] T. Borer and T. Davies. *Dirac video compression using open standards*. BBC R&D White Paper, WHP 117, September 2005.

[4] K. Onthriar, K.K. Loo and Z. Xue, *Performance Comparison of Emerging Dirac Video Codec with H.264/AVC*, International Conference on Digital Telecommunications, D.O.I. 10.1109/ICDT.2006.60.

[5] E. Dumic, M. Mustra and S. Grgic, *Comparison of Dirac and H.264/AVC Coding Quality using Objective Video Quality Measures*, 16th International Conference on Systems, Signals and Image Processing, IWSSIP, 2009, D.O.I. 10.1109/IWSSIP.2009.5367783.

[6] S. Kim and Y. S. Ho. *Rate control algorithm for H.264/AVC video coding standard based on rate-quantization model*, in: Proceedings of the IEEE International Conference on Multimedia and Expo, ICME’04, 1 June 2004, pp. 165-168.

[7] Y. K. Kim, Z. He and S. K. Mitra. *A novel linear source model and a unified rate control algorithm for H.263/MPEG-2/MPEG-4*, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP’01, 3 May 2001, pp. 17-80.

[8] F. L. P. Albini, *Geração e Avaliação de Artefatos em Vídeo Digital*, Dissertação de Mestrado, CPGEI / UTFPR, 2009.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. *Image quality assessment: from error visibility to structural similarity*. IEEE Trans. on Image Processing, vol. 13, issue 4, April 2004, pp. 600-612.

[10] Y. Shi et al. *Structure and Hue Similarity for Color Image Quality Assessment*. In International Conference on Electronic Computer Technology, February 2009, pp. 329-333.

[11] Z. Wang, H. R. Sheikh, and A. C. Bovik. *No-reference perceptual quality assessment of jpeg compressed images*. Proc. IEEE International Conference on Image Processing, September 2002.

[12] M. Abdel-Hady and R. Ward. *A Framework for Evaluating Video Transmission over Wireless Ad Hoc Networks*. In IEEE Conference on Communications, Computers and Signal Processing (PacRim), August 2007, pp. 78-81.

[13] Source videos, <http://media.xiph.org/video/derf/>.

[14] Matlab Image & Video Tools, <http://sprljan.com/nikola/matlab/yuv.html>.

[15] x264 Developers, <http://www.videolan.org/developers/x264.html>.

[16] H., Till. *Comparison of Open and Free Video Compression Systems*, <http://etill.net/projects/dirac_theora_evaluation/include/halbach-2009-dirac_theora-paper.pdf>.