An Adequate Objective Performance Evaluation of Dirac Video Codec Compared to H.264/AVC

André F. de Araujo
Communications Department
FEEC/UNICAMP
Campinas, Brazil
aaraujo@decom.fee.unicamp.br

Yuzo Iano
Communications Department
FEEC/UNICAMP
Campinas, Brazil
yuzo@decom.fee.unicamp.br

Rangel Arthur
School of technology
UNICAMP
Limeira, Brazil
rangel@ft.unicamp.br

Abstract— **Dirac, a relevant new open wavelet-based video codec, has been released recently. The main contribution of this paper is an adequate performance evaluation of this codec compared to the well-known H.264/AVC. This work differs significantly from previous ones which present unfair and inconclusive comparisons. Eight videos comprising three different resolutions with various degrees and kinds of motion are used to perform the simulations, as well as appropriate tuning of coding parameters for both codecs. In every case, H.264's objective visual quality results outperform Dirac ones, but some evidences indicate Dirac might provide better subjective quality.**

Keywords-Dirac video codec; H.264; Video Coding; Vide Compression; Wavelet-based video coding.

I. Introduction

Video coding is essential to enable video transmission and storage in modern communications systems. Most important video codecs, such as MPEG-based ones (and then H.264/AVC), employ the Discrete Cosine Transform (DCT) – or a similar transform based on it – as the transform responsible for eliminating spatial redundancy.

The Discrete Wavelet Transform (DWT) has gained importance recently, as it presents important advantages when compared to DCT, such as the possibility of working with the whole picture, instead of the need to partition into squared blocks. Moreover, subband decomposition results in more flexibility in terms of scalability in resolution and distortion.

Dirac is a recent open (royalty-free) video codec developed by the BBC R&D. Its architecture is similar to hybrid MPEG-based ones (as in [1]), except that the DWT is used for eliminating spatial redundancy. It was conceived to provide visual quality at H.264/AVC level.

Dirac creators explicitly argue that the codec was not designed to optimize objective visual quality measurements [2]. However, objective evaluations still provide important indices of the resulting quality and are able to indicate the general degree to which the compression is achieved.

Some works [3]-[6] have already attempted to provide an objective evaluation of Dirac against H.264, but these analyzes do not take into account important parameters so that the codecs can be compared adequately. These are then inconclusive.

In this work, Dirac is considered comparatively with regard to H.264/AVC, aiming to provide objective visual quality results for both codecs, so that a fair comparison is obtained. Initially, an overview of the main features of Dirac is outlined. Experimental results are derived using eight videos with three different resolutions and various degrees and kinds of motion with all relevant parameters adequately tuned.

II. Overview of Dirac Video Codec

Dirac is an open video codec developed by the BBC. It provides great compression efficiency (it is intended to compete with H.264/AVC) with a simple architecture, based on a small number of core tools. Its overall design is similar to a conventional hybrid motion-compensated codec, except that the function usually performed by a block transform in most standards (DCT, in most cases) is instead performed by the wavelet transform. Fig. 1 presents the overall architecture of the codec.

The encoder consists in four major modules, supporting both frames and fields: Transform and Quantization, Motion Estimation, Motion Compensation and Entropy Coding.

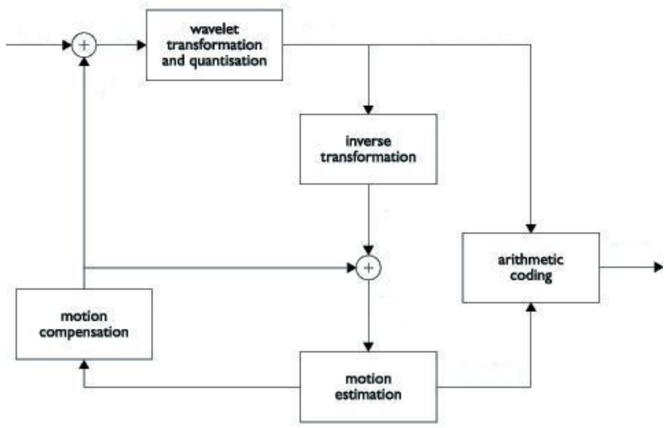

Figure 1. Overall architecture of Dirac Video Codec [7].

This work was supported by CAPES, RH-TVD/CAPES and FAEPEX/UNICAMP

A. Transform and Quantization

Transform operation is performed by wavelet filters. This key element allows for the support of any resolution without enlarging the codec toolset. A wide range of wavelet filters can be used. For compatibility with JPEG2000, Daubechies (9,7) filter is provided [8].

Both Intra and Inter pictures (a picture corresponding to a frame or a field) are wavelet transformed; Intra pictures are directly transformed, whilst Inter ones are motion compensated before entering this stage, resulting in a residual frame to be transformed.

The picture component data are coded in three stages. First, the data arrays are wavelet-transformed using separable wavelet filters and divided into subbands. Then, they are quantized (using RDO quantizes in the encoder). Finally, the quantized data is entropy coded. Fig. 2 illustrates this procedure.

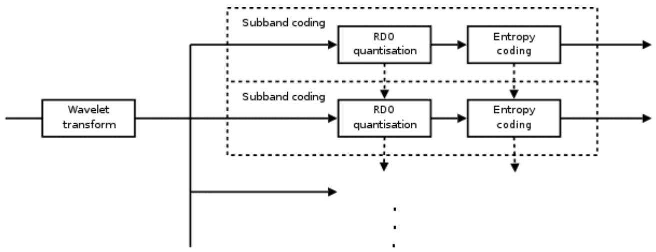

Figure 2. Stages involved in coding picture data [9].

Parent-child relationships between coefficients are also exploited when entropy coding wavelet data. This is based on the correlation between coefficients that represent the same area inside the picture.

Quantization follows by the use of a dead-zone quantizer, which is characterized by the zero interval twice as wide as the uniform one. This presents the advantage of simple implementation and good denoising results. The selection of quantizers is then performed by the minimization of the usual Lagrangian combination of rate and distortion:

$$J = D(Q) + \lambda R(Q) \tag{1}$$

where J is the expression to be minimized, D is the distortion, R the bit rate, Q the quantization factor and λ the parameter which determines the trade-off between rate and distortion.

B. Motion Estimation

This task is specific to the encoder, and represents the heaviest part of the encoding process. Dirac uses a 3-stage approach, using luminance samples only. Firstly, motion vectors are found for every block at pixel accuracy. In the second stage, these vectors are refined to sub-pixel accuracy (supporting until 1/8 pixel accuracy).

The metric used for block comparison in the first and second stages of the ME process is the usual Sum of Absolute Differences (SAD):

$$SAD = \sum_{j=0}^{N-1}\sum_{i=0}^{N-1} | c_{ij} - r_{ij} | \tag{2}$$

where c_{ij} and r_{ij} are the samples from the blocks of the reference and current figures and N is the block size.

Mode decision is the final step for the motion estimation process, and will choose between four possibilities of coding (Intra-coding, Inter-coding with reference 1, Inter-coding with reference 2 and Inter-coding with both references) and every possible partition choice.

Three partition levels are available: superblock (level 0), sub-superblock (level 1) and block (level 2) - they are shown in Fig. 3.

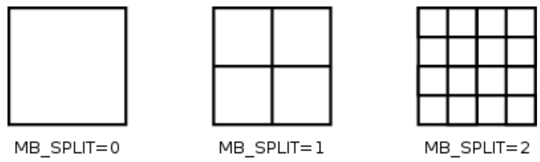

Figure 3. Partition levels available in Dirac [9].

In this step, the degree to which the motion vectors change is taken into account for the final vector decision, as the less they change, fewer bits are needed:

$$metric = SAD + \lambda \min\left[| V_x - pred_x | + | V_y - pred_y |, | V_x | + | V_y | \right] \tag{3}$$

where V refers to the motion vector to be evaluated, *pred* refers to the resulting median predicted vector and λ a parameter which weights the importance given to the second term. The factor $|V_x| + |V_y|$ is used to limit the excess of importance that might be assigned to the second term of the metric; when there is significant change in motion, for instance, the vector needs to be changed.

Dirac uses 3 types of picture. Intra pictures (I) are coded without reference to other pictures in the sequence. Level 1 pictures (L1) and Level 2 pictures (L2) are both inter pictures, i.e., they are coded with reference to other previously coded pictures. L1 pictures are forward-predicted only (they are also known as P-pictures) whereas L2 pictures are also known as B pictures (predicted from both earlier and later references).

C. Motion Compensation

Dirac uses Overlapped Block-based Motion Compensation (OBMC), which avoids block-edge artifacts that would be expensive to code using wavelets. Any block sizes can be used, with any degree of overlap selected. This information must be sent to the decoder. However, there must be an exact number of superblocks (4x4 set of blocks) horizontally and vertically. This is achieved by padding the data.

The OBMC method lies on a separable linear ramp mask, which acts as a weighed function on the predicting block. A pixel may fall within only one block or in up to four if it lies at the corner of a block.

D. Entropy Coding

Wavelet coefficient entropy coding in Dirac is based on three stages: binarization, context modeling and adaptive arithmetic coding. Fig. 4 illustrates this process.

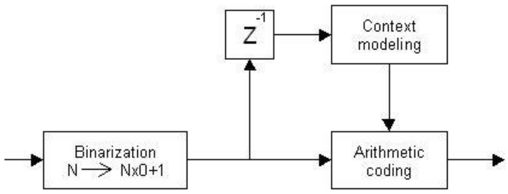

Figure 4. Entropy coding block diagram for wavelet coefficients [9].

A similar scheme is used for entropy coding of motion vectors, except that a prediction stage is employed before binarization and that different context models are used.

III. Performance Comparison

Eight video sequences were used to test the referred codecs, comprising different resolutions, degrees and kinds of motion. Their details are given in Table I, and Figs. 5-7 give sample frames of the videos, as well as Dirac-decoded ones with Hierarchical ME.

TABLE I. VIDEOS USED FOR THE EXPERIMENTS

Video	Resolution	Number of frames	Frame rate (fps)
canal352	352x288 (CIF)	68	12.5
snow352	352x288 (CIF)	75	12.5
squirrel352	352x288 (CIF)	75	12.5
canal720	720x576 (SDTV 576p)	50	25
snow720	720x576 (SDTV 576p)	50	25
squirrel720	720x576 (SDTV 576p)	50	25
snow1280	1280x720 (HDTV 720p)	30	60
squirrel1280	1280x720 (HDTV 720p)	30	60

The sequences are available online [10] in RGB format. They were converted to YUV 4:2:2 before compression. Each video was compressed several times with the same algorithm with different target bit rates, so that the performance can be analyzed correctly.

The encoders used were the reference encoders for both codecs implemented in C++. For Dirac codec, the 1.0.2 version was employed [11] while JM 15.1 was used for H.264 (main profile) [12]. The simulations were done using a PC equipped with Intel Pentium M 740, 1.73GHz, 768 MB RAM.

To provide a fair comparison between two codecs, some parameters must be tuned appropriately. The algorithm used for performing motion estimation (ME) affects evidently the results. Thus, both codecs were tested with Full Search ME with search range 32x32. For providing visual quality comparison resulting from the ME technique employed by Dirac, it was also tested with Hierarchical ME.

(a) (b)

Figure 5. (a) Sample frame from canal352 video (half resolution). (b) Reconstructed picture using Dirac with Hierarchical ME, with PSNR = 45.5379 and MSSIM = 0.992588.

(a) (b)

Figure 6. (a) Sample frame from snow352 video (half resolution). (b) Reconstructed picture using Dirac with Hierarchical ME, with PSNR = 49.2899 and MSSIM = 0.994792.

(a) (b)

Figure 7. (a) Sample frame from squirrel352 video (half resolution). (b) Reconstructed picture using Dirac with Hierarchical ME, with PSNR = 46.1565 and MSSIM = 0.995813.

The Group of Pictures (GOP) used also affects decisively the results. Then, for both codecs, the distance between P (L1) figures was defined to be 3, allowing the use of B (L2) figures – only the initial frame is of type I.

It is important to notice that none of the performance evaluations published so far [3]-[6] observes these restrictions, required to provide a fair comparison: different resolutions, degrees and kinds of motion, same ME algorithm and same GOP.

For providing visual quality comparison, two distortion metrics were used: the Peak Signal-to-Noise Ratio (PSNR) and the Mean Structural Similarity Index (MSSIM) [13].

Computational load is evaluated by the coding time per frame. Although the implementations here are not optimized, they are both research versions and can then provide a good idea of the computational complexity of each encoder.

A. Results

Due to the large quantity of visual quality results, only the most important ones are shown here. Figs. 8-13 give resulting PSNR x Bit rate and MSSIM x Bit rate curves, which show the visual quality achieved for the given bit rates. Table II summarizes main results concerning computational complexity: average encoding time per frame for each video and its percentage with respect to H.264 results. The results for Dirac with Hierarchical ME are not shown in Table II because they can't be compared, as different ME techniques are employed.

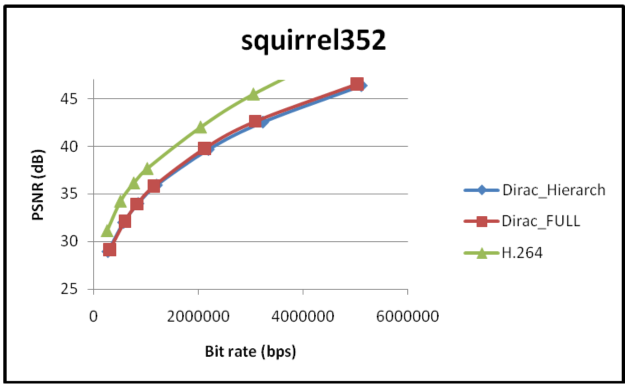

Figure 8. Visual quality results (PSNR x Bit rate) for squirrel352 video.

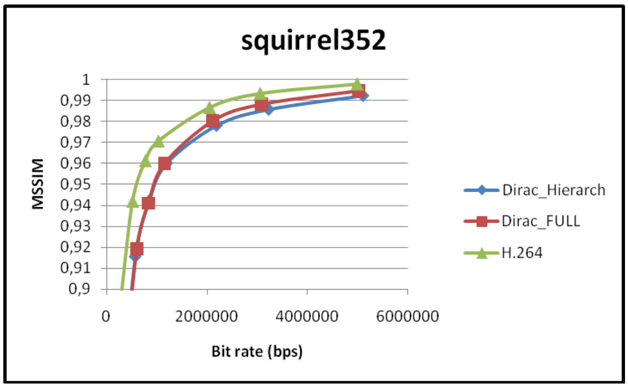

Figure 9. Visual quality results (MSSIM x Bit rate) for squirrel352 video.

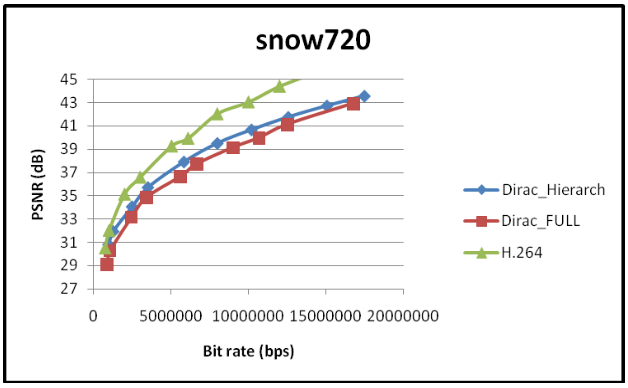

Figure 10. Visual quality results (PSNR x Bit rate) for snow720 video.

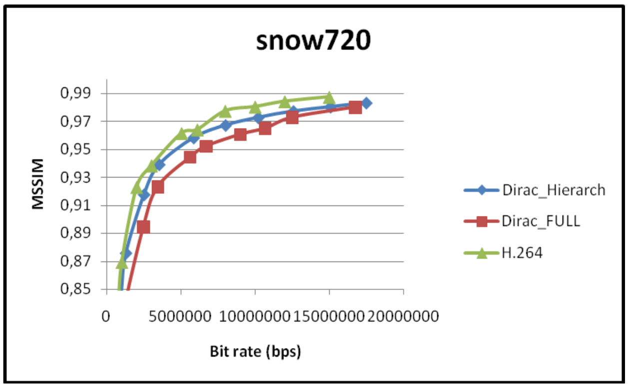

Figure 11. Visual quality results (MSSIM x Bit rate) for snow720 video.

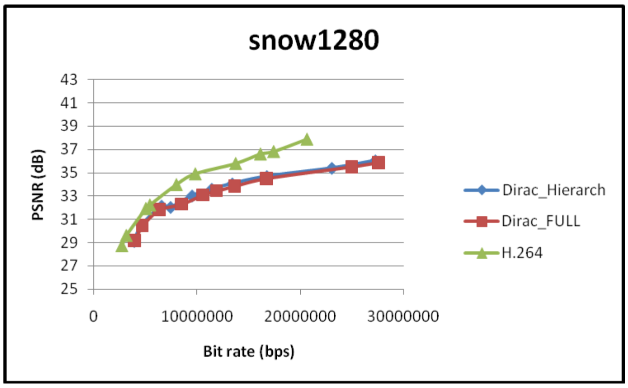

Figure 12. Visual quality results (PSNR x Bit rate) for snow1280 video.

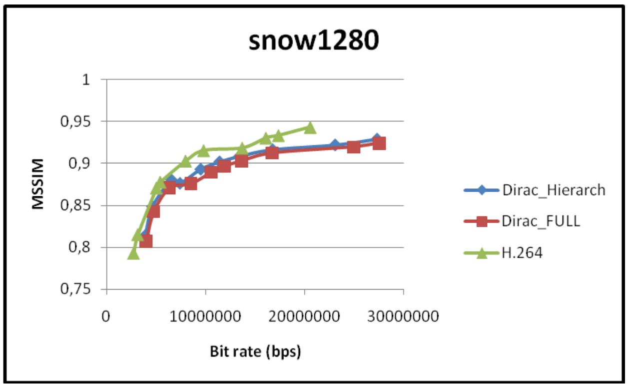

Figure 13. Visual quality results (MSSIM x Bit rate) for snow1280 video.

B. Discussion

It can be inferred that H.264 presents better objective visual quality, both in terms of PSNR and MSSIM. The difference between H.264's and Dirac's PSNR achieves up to 3dB, which is significant. Though, MSSIM results, by the slope of the results presented, do not confirm that the quality difference is that high.

Also, it can be noticed that Dirac with Hierarchical ME outperforms Dirac with Full Search in some cases. This can be explained by remarking that Hierarchical ME is able to search much further than the search range it is imposed to Full Search. For videos that contain large motion, such as snow720 and snow1280, this is very useful.

Regarding computational load, it is confirmed that Dirac algorithm is much simpler: 60% faster on average. It is much faster for every analyzed video, ranging from 44% to 83% speedup.

TABLE II. MAIN COMPUTATIONAL COMPLEXITY RESULTS FOR THE EXPERIMENTS

Video	H.264 - Avg. encoding time per frame (% comp. H.264)	Dirac - Avg. encoding time per frame (% comp. H.264)
canal352	38.08 (100)	20.93 (55)
snow352	38.98 (100)	21.78 (56)
squirrel352	39.73 (100)	20.89 (53)
canal720	148.01 (100)	45.32 (31)
snow720	153.06 (100)	73.23 (48)
squirrel720	157.24 (100)	51.67 (33)
snow1280	393.33 (100)	104.37 (27)
squirrel1280	412.57 (100)	71.66 (17)
Overall percentage average	***100%***	***40%***

As already stated, it is argued that Dirac was not conceived to optimize objective visual quality results. Analyzing some of the decoded pictures, it is indeed observed that Dirac presents better visual quality even if objective results indicate H.264 performs better. One example is given in Fig. 14: the body of the squirrel is much better defined in Dirac decoded frame. The image treated by H.264 is clearly much more blurred. Even in this case, PSNR and MSSIM are better for the image decoded with H.264.

(a) (b)

Figure 14. 10th frame of the squirrel1280 sequence (30% resolution). (a) Dirac-decoded picture with PSNR = 30.3596 and MSSIM = 0,876668. (b) H.264-decoded picture, with PSNR = 33.5563 and MSSIM = 0,929297.

IV. CONCLUSION

This work is motivated by the need to evaluate the recently-released Dirac video codec. As H.264/AVC is one of the best existing codecs, a comparison against the latter is provided. There are some works which try to perform this comparison [3]-[6], but none of them is concerned with tuning every appropriate parameter so that the results be relevant. The results derived by these works are then inconclusive.

Numerical results are derived from tests with 8 videos, comprising different degrees and kinds of motion and three different resolutions: CIF, SD 576p and HD 720p. Results clearly show that H.264 outperforms Dirac for the objective visual quality comparison. However, Dirac provides on average 60% complexity reduction.

Furthermore, it is observed that resulting visual quality for some frames is better for Dirac-treated ones, even if PSNR and MSSIM results indicate H.264-treated ones present better objective visual quality. Further work comprising proper subjective evaluations would be then very important.

Dirac is a relevant wavelet-based video codec which is still on its 1.0.2 release (the first official one was released on September 2008). It presents a real contribution in terms of wavelet-based codecs.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthr, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] T. Borer, T. Davies and A. Suraparaju, "Dirac video compression", BBC R&D White Paper, WHP 124, Sept. 2005.

[3] A. Ravi, "Performance analysis and comparison of Dirac video codec with H.264 / MPEG-4 Part 10 AVC", MSc. Thesis, Univesity of Texas at Arlington, Arlington, 2009.

[4] 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, Chalkida, 2009.

[5] T. Halbach, "Comparison of Open and Free Video Compression Systems", International Conference on Image Theory and Applications, 2009.

[6] K. Onthriar, K. K. Loo and Z. Xue, "Performance Comparison of Emerging Dirac Video Codec with H.264/AVC", International Conference on Digital Telecommunications, Cote D'Azur, 2006.

[7] BBC R&D, "The technology behind Dirac", available online at http://www.bbc.co.uk/rd/projects/dirac/technology.shtml. Accessed on May 9, 2010.

[8] BBC R&D, "Dirac Specification", v. 2.2.3, September 2008. Available online at http://diracvideo.org/ download/specification.

[9] BBC R&D, "The Dirac Algorithm", available online at the wiki of the project: http://diracvideo.org/wiki/index.php/Dirac_Algorithm. Accessed on May 9, 2010.

[10] BBC R&D, "Dirac Project". Available online at http://sourceforge.net/project/showfiles.php?group_id=102564&package_id=119507. Accessed on May 9, 2010.

[11] "Dirac Reference Software 1.0.2". Available online at http://sourceforge.net/project/showfiles.php?group_id=102564&package_id=112141. Accessed on May 9, 2010.

[12] "H.264/AVC JM15.1 Reference Software". Available online at http://iphome.hhi.de/suehring/tml. Accessed on Jun 9, 2009.

[13] 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, April 2004, vol. 13, no. 4, pp. 600-612.