PERFORMANCE COMPARISON OF VIDEO CODING STANDARDS USING LAGRANGIAN CODER CONTROL

Anthony Joch[*], *Faouzi Kossentini*[*], *Heiko Schwarz*[**], *Thomas Wiegand*[**], *Gary J. Sullivan*[***]

[*]University of British Columbia, [**]Heinrich Hertz Institute, [***]Microsoft Corporation

ABSTRACT

A unified approach to the coder control of video coding standards such as MPEG-2, H.263, MPEG-4, and the draft video coding standard JVT/H.26L/AVC is presented. Using this unified framework, the performance of the various standards is compared by means of PSNR and subjective testing results. The results indicate that JVT/H.26L/AVC compliant encoding can typically achieve essentially the same objective PSNR reproduction quality as encoders that are compliant with previous standards while requiring as little as 60% or less of the bit rate of the next best standard, particularly for higher-latency applications and particularly for more difficult source material. Subjective testing shows that the bit savings produced by this draft standard are even larger than the PSNR results indicate.

1. INTRODUCTION

The specifications of most video coding standards including MPEG-2 Video, H.263, MPEG-4 Visual, and JVT/H.26L/AVC[1] [6] describe only the bit-stream syntax and the decoding process in order to enable interoperability without imposing unnecessary constraints on implementation. Many coding parameters such as macroblock modes, motion vectors, and quantized transform coefficients have to be determined by the video encoder in a manner not defined in the standard. The chosen values determine the rate-distortion efficiency of the produced bit-stream of a given encoder.

In this paper, the operational control of MPEG-2, H.263, MPEG-4 and JVT encoders is optimized with respect to their rate-distortion efficiency using Lagrangian optimization techniques. The optimization is based on [1] and [2], where the encoder control for the ITU-T Recommendation H.263 is addressed. The Lagrangian coder control as described in this paper was integrated into compliant MPEG-2, H.263, MPEG-4 and JVT encoders. In addition to achieving performance gains, the use of similar rate-distortion optimization methods in all encoders allows a useful comparison between the encoders in terms of coding efficiency. Our experimental results based on PSNR measures indicate that equivalent quality can be achieved using JVT while requiring as little as 60% of the bit rate needed by the most advanced existing coding standards. Furthermore, subjective tests have shown that even larger bit rate savings can be achieved when perceptual quality is considered.

This paper is organized as follows. Section II gives an overview of the syntax features of MPEG-2 Video, H.263, MPEG-4 Visual, and JVT. The rate-distortion-optimized coder control is described in Section III, and experimental results are presented in Section IV.

2. OVERVIEW OF CODING ALGORITHMS

Although MPEG-2, H.263, MPEG-4, and JVT define conceptually similar coding algorithms, they contain features and enhancements that make them differ. These differences involve mainly the formation of the prediction signal, the block sizes used for transform coding, and the entropy coding methods. For descriptions of the MPEG-2, H.263, and MPEG-4 coding algorithms see [3], [4], [5]. Here we provide a brief overview of the JVT coding algorithm, since it has not been widely presented in the literature. For a more detailed description of H.26L, refer to the latest draft of the standard [6]. The description and experimental results presented herein refer to JVT Joint Working Draft 2 (JVT JWD2) [6].

The underlying coding scheme defined by JVT is superficially similar to that successfully employed in prior video coding standards, such as H.263 and MPEG-2. This includes the use of translational block-based motion compensation, block transformation, scalar quantization with an adjustable step size for bit rate control, zigzag scanning and run-length VLC coding of quantized transform coefficients. However, specific details within this structure and some key additional features differentiate JVT from all other standards.

The motion compensation model used in JVT is more flexible than those found in earlier standards. More specifically, JVT supports various rectangular partitions of each macroblock for motion-compensated coding, allowing greater flexibility than in MPEG-2, H.263 and

[1] It is intended that the draft standard produced by the Joint Video Team (JVT) will become ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10 or Advanced Video Coding (AVC)). We will henceforth call it JVT in this paper.

MPEG-4. Support for the use of multiple reference pictures for prediction is also included in the standard. Moreover, motion vectors can be specified with higher spatial accuracy than in earlier standards, with quarter-pixel accuracy as the default lower-complexity method and eighth-pixel accuracy available as a higher-complexity option. The conventional picture types known as I-, P-, and B-pictures are supported, with B-pictures more generalized than in earlier standards. The use of a powerful deblocking filter within the motion compensation loop is specified in order to reduce visual artifacts and improve prediction.

JVT is unique in that it employs a purely integer spatial transform which is primarily 4x4 in shape, as opposed to the usual floating-point 8x8 DCT specified with rounding-error tolerances as in earlier standards. The small size helps to reduce blocking and ringing artifacts, while the precise integer specification eliminates any mismatch between the encoder and decoder in the inverse transform. Extensive spatial prediction within frames is used for improved de-correlation in areas not using temporal prediction.

Two methods of entropy coding are supported in JVT. The first method, called Universal Variable Length Coding (UVLC), uses one single infinite-extent codeword set for all syntax elements. The coding efficiency can be improved if the more complex Context-Adaptive Binary Arithmetic Coding (CABAC) is used.

3. LAGRANGIAN VIDEO CODER CONTROL

The task of coder control is to determine a set of coding parameters, and thereby the bitstream, such that a certain rate-distortion trade-off is achieved for a given decoder. A particular emphasis is on Lagrangian bit-allocation techniques, which have emerged to form the most widely accepted approach in recent standard development, due to their effectiveness and simplicity.

For hybrid video coder control, the selection of motion vectors and the best coding mode for each macroblock can be optimized using Lagrangian minimization techniques. Let the Lagrange parameter λ_{MODE} and the quantizer value Q be given. The Lagrangian mode decision for a macroblock $\boldsymbol{S}_k$ proceeds by minimizing

$$J_{MODE}(\boldsymbol{S}_k, I_k \mid Q, \lambda_{MODE}) = D_{REC}(\boldsymbol{S}_k, I_k \mid Q) + \lambda_{MODE} R_{REC}(\boldsymbol{S}_k, I_k \mid Q) \quad (3)$$

where the macroblock mode I_k is varied over all possible coding modes available for a particular picture type and coding standard. The distortion $D_{REC}(\boldsymbol{S}_k, I_k|Q)$ and rate $R_{REC}(\boldsymbol{S}_k, I_k|Q)$ for the various modes represent the sum of the squared differences (SSD) between the reconstructed and original samples, and the number of bits required for encoding using a particular mode, respectively.

Motion-compensated modes require block motion estimation to select motion vectors $\boldsymbol{m}_i$ for block size $\boldsymbol{S}_i$ within a macroblock. The Lagrangian cost function

$$\boldsymbol{m}_i = \arg\min_{\boldsymbol{m} \in M} \{ D_{DFD}(\boldsymbol{S}_i, \boldsymbol{m}) + \lambda_{MOTION} R_{MOTION}(\boldsymbol{S}_i, \boldsymbol{m}) \} \quad (4)$$

is minimized over coding modes and motion vectors in the chosen search range. In our experiments, the sum of absolute differences (SAD) is used for D_{DFD}. The $R_{MOTION}(\boldsymbol{S}_i, \boldsymbol{m})$ is the number of bits to transmit all components of the motion vector $\boldsymbol{m}$. Experimental selection of Lagrangian multipliers is discussed in [7] [8].

4. COMPARISON OF STANDARDS

Two separate experiments were performed, each targeting a particular application area. The first experiment evaluates performance for video streaming while the second experiment targets video conferencing. The two applications are different in the sense that the delay constraints that are imposed in the video conferencing experiment are relaxed in the streaming case. The PSNR-based results described in the following two sub-sections indicate significant bit rate savings for JVT in each application. Perceptual testing indicates that the bit rate savings is effectively even larger.

4.1. Video Streaming/Distribution Applications

All encoders used only one I-picture at the beginning of a sequence, and 2 B-pictures have been inserted between each two successive P-pictures. Full search motion estimation with a range of 32 integer pixels was used by all encoders along with the Lagrangian Coder Control described in the previous section. The sequences used in this test consist of four QCIF sequences coded at 10 Hz and 15 Hz (Foreman, Container, News, Tempete) and four CIF sequences coded at 15 Hz and 30 Hz (Bus, Flower Garden, Mobile and Calendar, and Tempete).

The MPEG-2 Video encoder generated bit-streams that are compliant with the popular Main Profile at Main Level (MP@ML) and the H.263 encoder used the features of the High-Latency Profile (HLP). For MPEG-4 Visual, the Advanced Simple Profile (ASP) was used with the recommended de-blocking/de-ringing filter applied as a post-processing operation. For the JVT JM-2 coder, quarter-sample accurate motion compensation was used for QCIF sequences, and eighth-sample accurate motion compensation was used for CIF sequences, and entropy coding was performed using CABAC. We have generally used five reference frames for both H.263 and JVT with the exception of the News sequence, where we used a larger number of reference frames to exploit the unique redundancies (which can be detected by a compliant encoder) contained within this special sequence.

Table 1 presents the average bit-rate savings provided by each encoder relative to all other tested encoders over

the entire set of sequences and bit-rates. It can be seen that JVT significantly outperforms all other standards. On the most complex sequence of the test set, Mobile & Calendar (CIF, 30Hz), average bit-savings of more than 75% relative to MPEG-2 are realized. Bit-rate savings are as low as 50% on the Flower Garden sequence in CIF resolution (15Hz), with an average of 64% over the entire test set. JVT provides more than 35% bit-rate savings relative to MPEG-4 ASP and H.263 HLP.

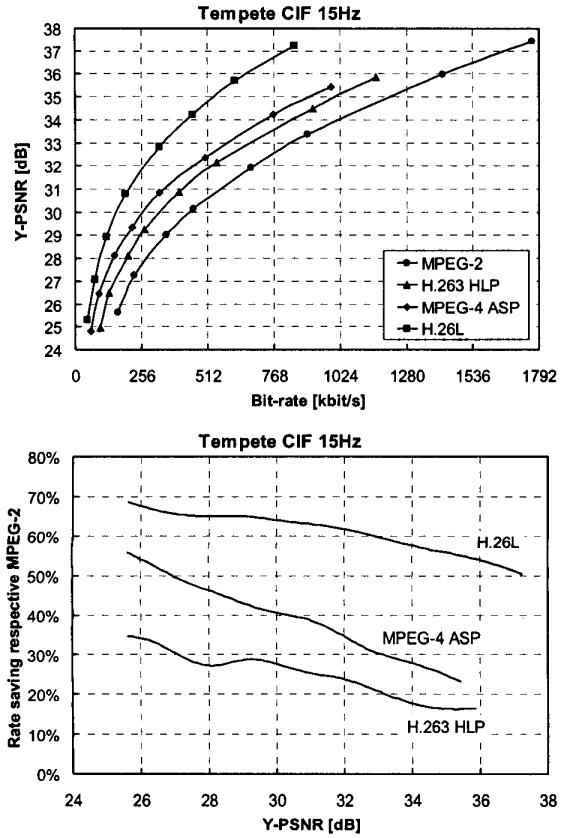

Figure 1: Sample rate-distortion and bit-rate savings curves for Tempete at 15 Hz in streaming comparison

	Average bit-rate savings relative to:		
Coder	MPEG-4 ASP	H.263 HLP	MPEG-2
JVT	39%	49%	64%
MPEG-4 ASP	-	17%	43%
H.263 HLP	-	-	31%

Table 1: Average bit rate savings for video streaming

4.2. Video Conferencing Applications

This experiment evaluates coding performance for interactive video applications, such as videoconferencing, in which low delay and real-time encoding capability are the key requirements. Such applications generally support low to medium bit-rates and picture resolutions, with QCIF resolution at 10-128 kbits/s and CIF resolution at 128-512 Kbits/s being the most common. A set of four QCIF sequences encoded at 10Hz and 15Hz and four CIF sequences encoded at 15Hz and 30Hz that represent a variety of conversational content were used in this experiment. The QCIF sequences are: Akiyo, Foreman, Mother and Daughter, and Silent Voice. The CIF sequences are: Carphone, Foreman, Paris, and Sean.

Encoders that are included in this comparison are compliant with the following standards/profiles: the H.263 Baseline and Conversational High Compression (CHC) Profiles, the MPEG-4 Simple Profile (SP) and ASP, and JVT. Since profiles are not yet finalized for JVT, the corresponding encoder is configured to provide the best possible rate distortion performance subject to the low-delay constraints imposed for this experiment, including CABAC entropy coding. Both the H.263 CHC and JVT encoders used five reference pictures for long-term prediction.

In all bitstreams, only the first picture was intra coded, with all of the subsequent pictures being temporally predicted (P-pictures). A motion search range of 32 integer pixels was employed by all encoders with the exception of H.263 Baseline, which is constrained by its syntax to a maximum range of 16 integer pixels.

In order to satisfy the low delay and complexity requirements of interactive video applications, this test did not include the use of B-pictures for any design because of the strict delay constraints of interactive applications. The global motion compensation feature of MPEG-4 ASP was also not used. Therefore, the only significant difference in terms of rate-distortion performance between the MPEG-4 SP and the ASP results in this experiment is that the ASP uses quarter-pixel accurate motion compensation, whereas the SP uses only half-pixel accuracy.

As in the first experiment, we present both rate-distortion curves for luminance component, as well as plots of bit-rate savings relative to the poorest performing encoder. As should be expected, it is the H.263 Baseline encoder that provides the worst performance, and it therefore serves as the basis for comparison. Figure 2 shows the rate-distortion plots and the bit-rate saving plots for three selected test sequences. The average bit-rate savings results over the entire test set are given in Table 2.

It is immediately clear from these results that JVT outperforms all of the other standards by a substantial margin. Bit-rate savings of more than 40% relative to H.263 Baseline are realized. Relative to MPEG-4 ASP and H.263 CHC, JVT provides more than 25% bit-rate savings. These reported bit-rate savings are lower than was measured in the first experiment for video streaming applications. This is related to the choice of typical videoconferencing sequences for the second experiment. These sequences are generally characterized by low or

medium motion as well as low spatial detail. The largest improvements of coding efficiency for JVT are obtained for complex sequences such as Mobile & Calendar.

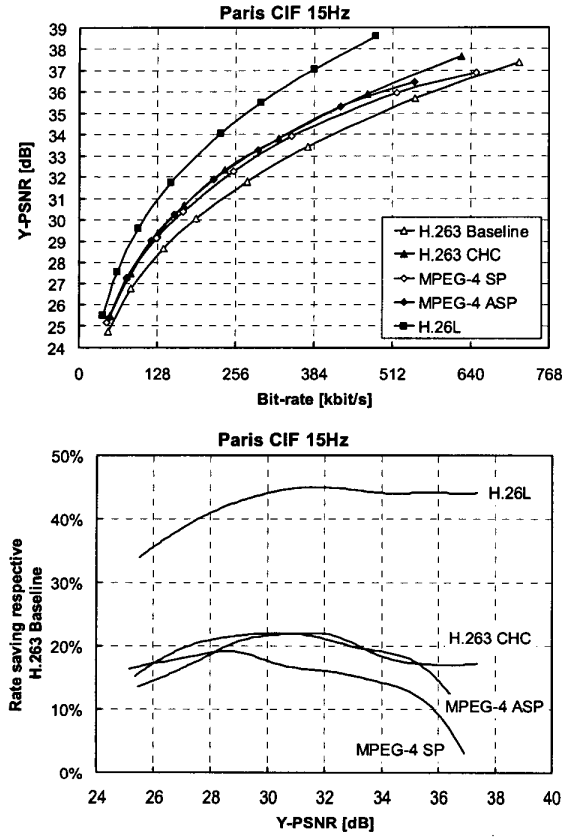

Figure 2: Rate-distortion and bit-rate savings curves for Paris at 15 Hz in conversational video comparison

Coder	Average bit-rate savings relative to:			
	MPEG-4 ASP	H.263 CHC	MPEG-4 SP	H.263 Base
JVT	28%	32%	34%	45%
MPEG-4 ASP	-	7%	10%	24%
H.263 CHC	-	-	2%	18%
MPEG-4 SP	-	-	-	16%

Table 2: Average bit-rate savings for conversation use

4.3. Perceptual Comparison Results

We have conducted informal subjective testing of the sequences generated with the various the rate-distortion optimized encoders. The purpose of these tests is to establish whether the PSNR-based results presented in the previous sub-sections provide an adequate measure of the bit rate savings that can be achieved with JVT while maintaining equivalent perceived quality. These tests have shown that when sequences of equivalent PSNR are presented, viewers tend to prefer those encoded with JVT to those of other standards. These results indicate that when perceptual quality is taken into account, JVT can provide even larger bit rate savings. The smaller block sizes used for transform coding and motion compensation, in combination with the powerful in-loop deblocking filter likely play an important role in the improved subjective quality. Further study is needed to quantify the perceptual benefit identified in these preliminary tests.

5. CONCLUSIONS

The impressive performance of JVT compliant encoders clearly demonstrates the potential importance of this standard in future video applications. Although JVT coding shares the common hybrid video coding structure with previous standards, it provides added features and increases flexibility, which enables improved coding efficiency for potentially increased complexity at the encoder. Further study is needed to track the performance impact of future changes as the standard matures to final form, to study expected performance in higher-quality "entertainment" applications, as well as to quantify the perceptual aspects of performance.

6. REFERENCES

[1] Thomas Wiegand and Barry D. Andrews: "An Improved H.263 Coder Using Rate-Distortion Optimization," ITU-T/SG16/Q15-D-13, April 1998, Tampere, Finland.

[2] M. Gallant, G. Côté, and F. Kossentini, "Description of and Results for Rate-Distortion Based Coder," ITU-T/SG16/Q15-D-47, April 1998, Tampere, Finland.

[3] B.G. Haskell, A. Puri, A.N. Netravalli, Digital Video: An Introduction to MPEG-2, Chapman and Hall, NewYork, USA, 1997.

[4] G. Côté, B. Erol, M. Gallant, and F. Kossentini. "H.263+: Video Coding at Low Bit Rates", In IEEE Transactions on Circuits and Systems for Video Technology, 8(7):849–866, November 1998.

[5] ISO/IEC JTC1, "Coding of audio-visual objects – Part 2: Visual," ISO/IEC 14496-2 (MPEG-4 visual version 1), April 1999; Amendment 1 (version 2), February, 2000; Amendment 4 (streaming profile), January, 2001.

[6] T. Wiegand (ed.), "Working Draft Number 2, Revision 2 (WD-2)," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-B118r2, February 2002.

[7] G. J. Sullivan and T. Wiegand "Rate-Distortion Optimization for Video Compression," in IEEE Signal Processing Magazine, vol. 15, no. 6, pp. 74-90, Nov. 1998.

[8] T. Wiegand and B. Girod, "Lagrangian Multiplier Selection in Hybrid Video Coder Control," in Proc. ICIP 2001, Thessaloniki, Greece, October 2001.