A New Rate Control Scheme for H.264/AVC

Yongdong Zhang
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
zhyd@ict.ac.cn

Wu Yuan
Graduate University of Chinese Academy of Sciences, Beijing 100080, China
wyuan@ict.ac.cn

Shouxun Lin
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
sxlin@ict.ac.cn

Abstract

A novel achievement to optimum bit allocation for H.264/AVC is presented, which is different from JVT-O016 by using a simple linear rate instead of the well-known MPEG-4 Q2 model. To find the global optimum, we resort to Lagrange Optimization technique and develop a close-form formula to the optimum problem. It is shown via extensive experiments that the new rate control (RC) scheme exceeds JVT-G012, the current standardized RC scheme and is comparative to JVT-O016 in coding performance.

1. Introduction

For rate control, its main task is to regulate DCT coefficients quantization to adapt to the actual channel bandwidth, and meanwhile achieve high-quality reconstructed pictures at the decoder. Several rate control schemes have been reported in the literature, including TMN-8[9], TM-5[7] and VM-18 [8], etc.

Generally speaking, a typical rate control scheme can be decomposed into two steps, which are bit allocation and subsequent achievement of the bits target. The insight behind optimal bit allocation and accurate achievement is to precisely approximate rate-distortion (R-D) behaviors of video content via mathematical modeling. Because video coding is a highly nonlinear process, it is a difficult task to precisely approximate R-D behaviors by using a close-form formula, and an empirical approach is most often used indeed[5]. Based on R-D models, Lagrangian optimization or linear programming methods are employed to achieve optimum bit allocation.

As a new generation of video coding standards, H.264/AVC[6] greatly outperforms prior coding standards in coding performance by exploiting lots of complicated coding methods. As a side-effect, H.264/AVC gets highly complicated. The introduction of Lagrangian coder control method into a H.264/AVC-complainted coder greatly challenges the design of rate control schemes. The Lagrangian coder control method demands quantization parameter as a prerequisite to the RDO for inter/intra prediction, and consequently couples inter/intra prediction and rate control much tightly. Since accurate R-D modeling on the residual can only be conducted after inter/intra prediction, the peculiarity of Lagrangian coder control method leads to the chicken and egg dilemma. Thus, the design of rate control scheme for H.264/AVC is quite different from prior standards, and conventional rate control scheme cannot be directly employed in a H.264/AVC-complianted coder.

Several works related to rate control have contributed to H.264/AVC, such as JVT-F086 [3], JVT-G012[1] and JVT-O016[4][11]. In JVT-F086, a multiple-pass scheme is presented to circumvent the dilemma, while in JVT-G012 and JVT-O016, a linear MAD model is used indeed. In JVT-G012, the conventional MPEG-4 Q2 model is employed to calculate the quantization parameter. Compared with JVT-G012, JVT-O016 improves the accuracy of MPEG-4 Q2 and MAD models by jointly considering spatial and temporal correlations. Moreover, JVT-O016 achieves optimum bit allocation at a macroblock-by-macroblock basis in a TMN-8-alike way, and therefore significantly improves the coding performance.

In this paper, an approach similar to JVT-O016 is presented with difference in using a linear rate model instead of a quadratic rate model and a quadratic distortion model instead of a linear distortion model.

0-7695-2650-0/06/$20.00 (c) IEEE
Authorized licensed use limited to: INSTITUTE OF COMPUTING TECHNOLOGY CAS. Downloaded on January 13, 2009 at 01:52 from IEEE Xplore. Restrictions apply.

Both of the two schemes focus on rate control for P-frames.

The paper is organized as follows. First, in Section II, we derive R-D models. And then, we deduce a close-form solution to optimum bit allocation in Section III. We propose a rate-distortion optimized rate control scheme in Section IV. Extensive experiments are conducted to evaluate the performance of the proposed rate control scheme in Section V. This paper concludes with Section VI.

2. Rate and Distortion Modeling

As is known, the DCT coefficients of the difference frame are approximately uncorrelated and *Laplacian* distributed. The entropy H of the quantized residue DCT coefficients is given by [10].

$$H(\alpha)=-\left(1-\exp\left(-2/\alpha\right)\right)\log_2\left(1-\exp\left(-2/\alpha\right)\right)-2\sinh\left(2/\alpha\right)$$
$$\left(\log_2\left(\sinh\left(2/\alpha\right)\right)\frac{1}{\exp(1/\alpha)-1}-\frac{1}{\alpha\ln 2}\frac{\exp(1/\alpha)}{\left(\exp(1/\alpha)-1\right)^2}\right)$$
(1)

where $\alpha=\dfrac{MAD}{Q}$, Q stands for the quantizer step, and MAD stands for mean absolute difference.

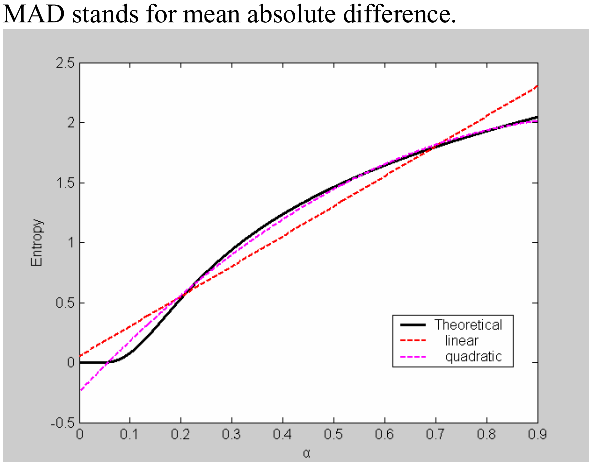

Fig. 1. the black solid line is the entropy of a Q-quantized *Laplacian* distribution with respect to α. The red dashed line is the linear fitting curve while the magenta is the quadratic one.

We plot $H(\alpha)$ in Fig.1, and we can see that $H(\alpha)$ monotonically increases with α. Since (1) is too complicated to use in practice, we want to find a formula in a simple form for a substitute. We plot its quadratic and linear fits in Fig.1. It is shown that the quadratic curve is more fitting than the linear one. In JVT O016, the quadratic fitting is selected to approximate R-Q relationship. However, it is worthwhile to note that existing an extremum on a quadratic curve breaks down the being that H vary monotonically with respect to Q or α. Moreover, additional Taylor expansion on MPEG-4 Q2 model increases computational complexity and even decreases the accuracy of the finial quadratic rate model especially in the case of coarse granular quantization. In this work, a simple linear rate model is proposed as the follows.

$$R=A\alpha+B=A\frac{MAD}{Q}+B \tag{2}$$

where A and B stand for model parameters and can be calculated using statistical linear regression analysis. We follow the same way and criteria to predict MAD, and select data points for the regression analysis as what JVT-O016 does. In JVT-O016, a linear model is used to predict MAD.

$$MAD_{i,j}^{K}=\rho MAD_{i,j}^{K-1}+\gamma \tag{3}$$

It is easy to see that equation (2) increases monotonically with regard to α, and decreases monotonically with regard to Q. Let $MB_{i,j}^{K}$ denote the macroblock at the j^{th} row and the i^{th} column in the K^{th} P frame. Following the way in JVT-O016 to predict overhead rate H, we can obtain a formula to estimate the total bit rate for $MB_{i,j}^{K}$:

$$R_{i,j}^{K}(Q)=A_{i,j}^{K}\frac{MAD_{i,j}^{K}}{Q_{i,j}^{N}}+B_{i,j}^{K}+H_{i,j}^{K-1} \tag{4}$$

Now, let's turn to the distortion. The distortion is introduced by quantizing the residue DCT coefficients. In practice, several statistics tools such as sum of absolute differences or its equivalents known as mean absolute differences, sum of squared differences or its equivalents known as mean squared differences or peak signal-to-noise ratio (PSNR) are used to evaluate the distortion. Mean absolute differences are used as the distortion measurement, and result in a linear distortion model in JVT-O016. Considering the fact that PSNR is most often employed as the objective measure of picture quality, we adopt mean squared differences in this work.

We assume that distortion is uniformity distributed, and the variance of $MB_{i,j}^{K}$ can be calculated as the follow.

$$D_{i,j}^{K}=\frac{(Q_{i,j}^{K})^2}{12} \tag{5}$$

Considering that correlations exist among temporally neighboring macroblocks, equation (5) is updated as

Authorized licensed use limited to: INSTITUTE OF COMPUTING TECHNOLOGY CAS. Downloaded on January 13, 2009 at 01:52 from IEEE Xplore. Restrictions apply.

$$D_{i,j}^K = \chi_{i,j}^K \frac{(Q_{i,j}^K)^2}{12} \quad \text{where} \quad \chi_{i,j}^K = \frac{D_{i,j}^{K-1}}{Q_{i,j}^{K-1}} \tag{6}$$

3. Optimum Bit Allocation

Based on equation (4) and (6), we derive a formula for the optimal quantizer step $Q_{i,j}^{K*}$ to minimize the total distortion of the K^{th} P-frame under the channel bandwidth constraint. This kind of optimization problem can be converted into a mathematic problem like the following.

$$\begin{aligned} &\min_{Q_{1,1}^K,\cdots,Q_{M,N}^K} \sum_{i=1,j=1}^{M,N} \chi_{i,j}^K \frac{(Q_{i,j}^K)^2}{12} \\ &s.t. \quad \sum_{i=1,j=1}^{M,N} \left(A_{i,j}^K \frac{MAD_{i,j}^K}{Q_{i,j}^K} + B_{i,j}^K + H_{i,j}^{K-1} \right) = R_{frame} \end{aligned} \tag{7}$$

where R_{frame} is the target bits used to code the K^{th} P-frame. By applying Lagrangian optimization method, the constrained minimization problem of (7) can be translated to the following unconstrained problem:

$$\begin{aligned} &\min_{Q_{1,1}^K,\cdots,Q_{M,N}^K} \sum_{i=1,j=1}^{M,N} \chi_{i,j}^K \frac{(Q_{i,j}^K)^2}{12} + \\ &\lambda \left(\sum_{i=1,j=1}^{M,N} \left(A_{i,j}^K \frac{MAD_{i,j}^K}{Q_{i,j}^K} + B_{i,j}^K + H_{i,j}^{K-1} \right) - T_{frame} \right) \end{aligned} \tag{8}$$

After some straightforward manipulations, we obtain the optimal quantizer step:

$$Q_{i,j}^{K*} = \frac{\sum_{m=1,n=1}^{M,N} \sqrt[3]{(A_{m,n}^K MAD_{m,n}^K)^2 \chi_{m,n}^K}}{T_{frame} - \sum_{m=1,n=1}^{M,N} (B_{m,n}^K + H_{m,n}^{K-1})} \sqrt[3]{\frac{A_{i,j}^K MAD_{i,j}^K}{\chi_{i,j}^K}} \tag{9}$$

After then, the corresponding quantizer parameter $QP_{i,j}^{K*}$ can be calculated from the quantizer step $Q_{i,j}^{K*}$.

4. Rate Control

We follow JVT-G012 to allocate the frame-level target T_{frame} for every P-frame except the first one in a GOP. In the literature of JVT-G012, both I-frame and the first P-frame are encoded by using a single quantizer parameter, and our efforts are focused on rate control process at a macroblock-by-macroblock basis for the rest P-frames. In the following, a brief description is given.

Step 1: *Initialization*

Set i=0, j=0.

Step 2: *Compute $Q_{i,j}^{K*}$ for $MB_{i,j}^K$.*

Step 2.1: Set ii=i, jj=j, ST=0, and SU=0; and then do the following iteration.

Step 2.2: Use linear regress method to calculate the MAD model parameters $\rho_{ii,jj}^K$ and $\gamma_{ii,jj}^K$, and then evaluate $MAD_{ii,jj}^K = \rho_{ii,jj}^K MAD_{ii,jj}^K + \gamma_{ii,jj}^K$. If $MAD_{ii,jj}^K$ is negative, then it is modified to 0.

Step 2.3: Use linear regress method to calculate the rate model parameters $A_{ii,jj}^K$ and $B_{i,j}^K$.

Step 2.4: Compute $\chi_{ii,jj}^K$ according to (6).

Step 2.5: *Loop Condition*—If the macroblock right to $MB_{ii,jj}^K$ is available, set ii=ii+1; otherwise set ii=0, jj=jj+1. If all macroblocks have been processed, then jumps to *Step* 3, otherwise jumps to *Step* 2.2.

Step 3: If $T_{frame} - ST$ is below zero, then we set the current QP = the *previous* $QP + 1$, and jumps to *Step* 6.

Step 4: Otherwise compute $Q_{i,j}^K = \dfrac{SU}{T_{frame} - ST} \sqrt[3]{\dfrac{A_{i,j}^K MAD_{i,j}^K}{\chi_{i,j}^K}}$ and derive the current QP from $Q_{i,j}^K$.

Step 5: Adjust QP by that the absolute difference between the current QP and the previous QP doesn't exceed 1, in order to maintain the smoothness of visual quality.

Step 6: *Macroblock Encoding.*

Step 7: *Post-Encoding*

Record current coding data, including the overhead bits count $H_{i,j}^K$, the distortion ratio $\chi_{i,j}^K = \dfrac{D_{i,j}^K}{(actualQ_{i,j}^K)^2}$, $\dfrac{MAD_{i,j}^K}{actualQ_{i,j}^K}$ (where $actualQ_{i,j}^K$ is the actual quantizer step used for $MB_{i,j}^K$,) $MAD_{i,j}^K$, and the texture bits count $R_{i,j}^K$.

If the macroblock right to $MB_{i,j}^K$ is available, set i=i+1; otherwise set i=0, j=j+1. If all macroblocks have been processed, then proceeds to the subsequent frame, otherwise jumps to *Step* 2.

5. Experimental Results

Authorized licensed use limited to: INSTITUTE OF COMPUTING TECHNOLOGY CAS. Downloaded on January 13, 2009 at 01:52 from IEEE Xplore. Restrictions apply.

The proposed rate control scheme has been implemented based on JM-9.4[2] coder. In our experiments, the well-known sequences of ‘Akoyi’, ‘Claire’, ‘Carphone’, ‘Coastguard’, ‘Football’, ‘News’, ‘Paris’ and ‘Suzie’ are used, where all are in format of QCIF(4:2:0) except that Football is in format of CIF(4:2:0). Test conditions are listed in Table-1. The initial QP is set to 43 for ‘Carphone’, 37 for ‘Football’, and 35 for all the rest.

Table 1. Test Conditions

MV resolution	**1/4 pel**	*Reference Frames*	**5**
Hadamard	**ON**	*Symbol Mode*	**CABAC**
RD optimization	**ON**	*GOP structure*	**IPPP**
Search Range	**±32**	*IntraPeriod*	**0**
Restrict Search Range	**2**	*Frames to be coded*	**100***

Peak-signal-to-noise-ratio (PSNR) and average bit prediction error (BPE) comparisons among results achieved by our proposed RC scheme, JVT-G012 and JVT-O016 are presented in Fig 2-5 and Table-2. BPE is defined as $|T_{actual} - T_{frame}| / T_{frame} \times 100\%$ where T_{actual} and T_{frame} are the actual and target bits count. From Fig 2 and Table-2 , it can be seen that compared to the JVT-G012 rate control scheme both the proposed scheme and JVT-O016 don’t only yield a much smaller prediction error but also achieves better image quality. The improvements with gains up to 1.21 dB in PSNR per frame over JVT-G012 can be achieved by the proposed RC scheme. The experimental results also show that the proposed RC scheme is comparable to JVT-O016 in coding performance.

6. Conclusion

In this work, a linear rate model is proposed to approximate the R-Q relationship of DCT coefficients. It varies monotonously in regard of QP, so doesn’t suffers what the MPEG-4 Q2 model does due to it’s extremum which breaks down the being that the bit rate varies monotonically with respect to *QP*. The optimum bit allocation is a little simpler than JVT-O016, while achieves a similar coding performance. The computational complexity can be reduced by using a linear distortion model instead of the quadratic one.

* Note, ‘Football’ consists of 88 frames.

7. Acknowledgment

This work was supported jointly by the National Nature Science Foundation of China (60302028) and the Key Project of International Science and Technology Cooperation (2005DFA11060).

8. References

[1] Z.G. Li, F. Pan, K.P. Lim, G.N. Feng, X. Lin and S. Rahardaj, “Adaptive basic unit layer rate control for JVT,” JVT-G012, 7th meeting, Pattaya II, Thailand, 7-14, March, 2003.

[2] JM9.4, http://iphome.hhi.de/suehring/tml/download/old_jm/jm94.zip.

[3] S.W. Ma, W. Gao, Y. Lu and H. Q. Lu, “Proposed draft description of rate control on JVT standard,” JVT-F086, 6th meeting, Awaji, Japan, 5-13, December, 2002.

[4] W. Yuan, S.X. Lin, Y.D. Zhang, W. Yuan and H.Y. Luo, **“**Optimum Bit Allocation and Rate Control for H.264/AVC,” JVT-O016, 15th Meeting: Busan, KR, 16-22 April, 2005.

[5] Zhihai He and Sanjit K. Mitra, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via p-domain Source Modeling," *IEEE Transactions on Circuits and System for Video technology*, October 2002.

[6] T. Wiegand and G. J. Sullivan, “Draft ITU-T Recommendation H.264 and Final Draft International Standard of Joint Video Specification (ITU-T Recommendation H.264 | ISO/IEC 14496-10 AVC)”, Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G050, Pattaya, Thailand, Mar. 2003.

[7] MPEG-2, Test Model 5, Doc. ISO/IEC JTC1/SC29 WG11/93-225b. Test Model Editing Committee, Apr. 1993.

[8] Video Group, “Text of ISO/IEC 14496-2 MPEG-4 Video VM-Version 8.0,” ISO/IEC JTC1/SC29/WG11 *Coding of Moving Pictures and Associated Audio MPEG 97/W1796*, Stochholm, Sweden, July 1997.

[9] J. Ribas-Corbera and S. Lei, “Rate control in DCT video coding for low-delay communications,” *IEEE Trans. Circuits Syst. Video Technol.*,vol. 9, pp. 172–185, Feb. 1999.

[10] F. Moscheni, F. Dufaux and H. Nicolas, “Entropy criterion for optimal bit allocation between motion and prediction error information”, *Proc. SPIE Visual Commun. And Image Proc.*, pp. 235-242, Nov. 93

0-7695-2650-0/06/$20.00 (c) IEEE

Authorized licensed use limited to: INSTITUTE OF COMPUTING TECHNOLOGY CAS. Downloaded on January 13, 2009 at 01:52 from IEEE Xplore. Restrictions apply.

[11] W. Yuan, S.X. Lin, Y.D. Zhang, W. Yuan and H.Y. Luo, "*Optimum Bit Allocation and Rate Control for H.264/AVC*," to appear in *IEEE Trans. Circuits Syst. Video Techno. vol.16, Jun. 2006.*

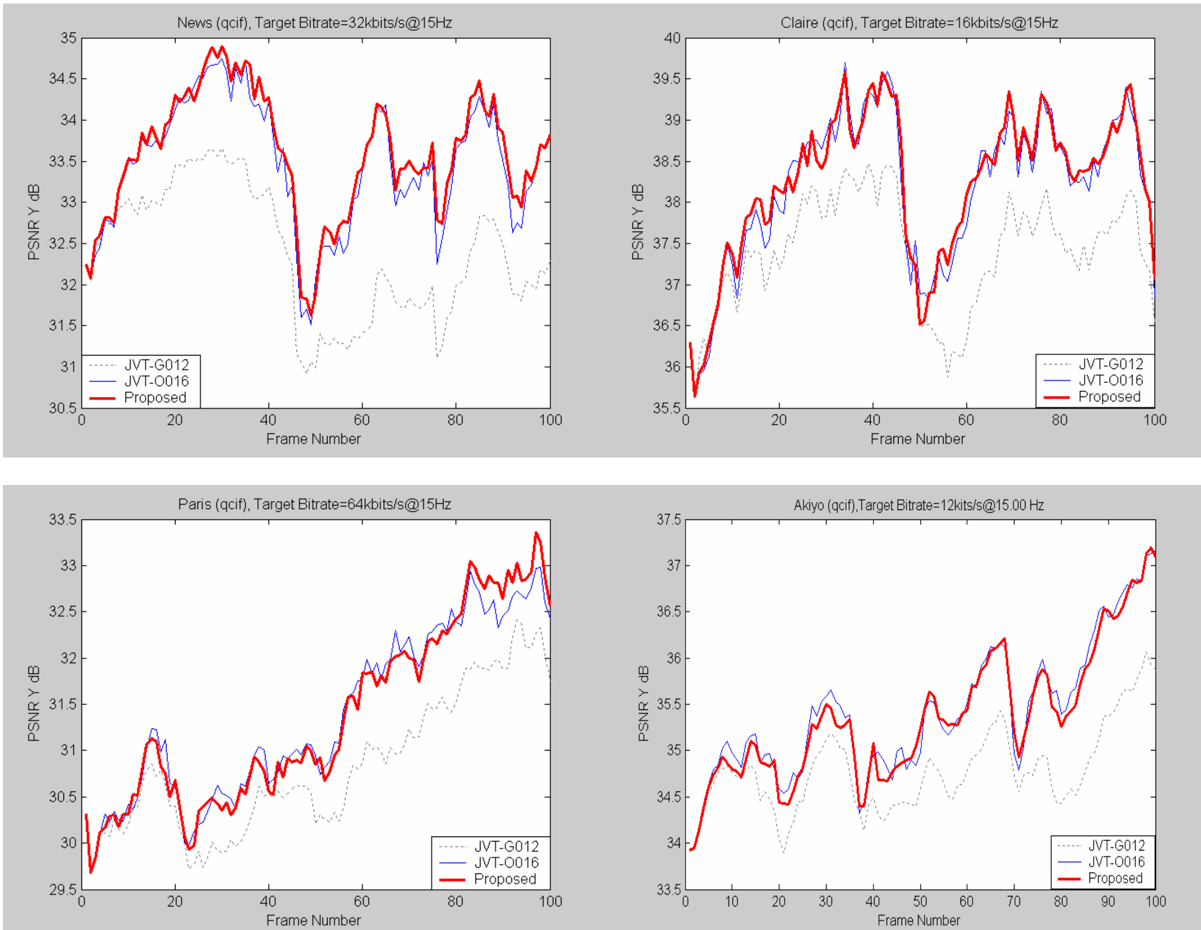

Fig.2. Comparison of PSNR for each frame with our proposed RC scheme with JVT-G012 and JVT-O016

0-7695-2650-0/06/$20.00 (c) IEEE
Authorized licensed use limited to: INSTITUTE OF COMPUTING TECHNOLOGY CAS. Downloaded on January 13, 2009 at 01:52 from IEEE Xplore. Restrictions apply.

TABLE 2 **Comparison Results. (Red: improvement, Blue: deterioration)**

Sequences	Encoder	SNR Y	SNR U	SNR V	BPE	Rate (kbit/s)	Sequences	Encoder	SNR Y	SNR U	SNR V	BPE	Rate (kbit/s)
Akiyo	**Proposed**	**35.36**	**38.28**	**40.06**	**24.01%**	**12.06**	**Carphone**	**Proposed**	**28.65**	**36.21**	**36.76**	**13.28%**	**12.03**
	JVT-G012	34.78	38.48	39.95	65.01%	12.07		**JVT-G012**	28.26	36.04	36.53	83.68%	12.1
	Gains	0.58	-0.2	0.11	-41.00%	-0.01		***Gains***	0.39	0.17	0.23	-70.40%	-0.07
	JVT-O016	35.42	38.31	40.08	22.40%	12.08		**JVT-O016**	28.59	36.05	36.63	11.35%	12.03
	Compare	-0.06	-0.03	-0.02	1.61%	-0.02		***Compare***	0.06	0.16	0.13	1.93%	0
Claire	**Proposed**	**38.23**	**38.53**	**40.42**	**21.14%**	**16.16**	**Football**	**Proposed**	**33.99**	**38.73**	**40.8**	**2.48%**	**512.13**
	JVT-G012	37.39	37.81	40.56	92.11%	16.32		**JVT-G012**	33.79	38.8	40.82	10.27%	513.89
	Gains	0.84	0.72	-0.14	-70.97%	-0.16		***Gains***	0.2	-0.07	-0.02	-7.79%	-1.76
	JVT-O016	38.18	38.48	40.55	18.57%	16.13		**JVT-O016**	33.91	38.78	40.83	3.13%	512.24
	Compare	0.05	0.05	-0.13	2.57%	0.03		***Compare***	0.08	-0.05	-0.03	-0.65%	-0.11
Coastguard	**Proposed**	**27.43**	**39.84**	**40.9**	**6.88%**	**24.05**	**News**	**Proposed**	**33.58**	**37.63**	**38.31**	**7.16%**	**32.04**
	Proposed	27.35	40.02	40.87	34.71%	24.12		**JVT-G012**	32.37	37.62	38.3	22.45%	32.06
	Gains	0.08	-0.18	0.03	-27.83%	-0.07		***Gains***	1.21	0.01	0.01	-15.29%	-0.02
	JVT-O016	27.38	39.89	40.91	5.80%	24.06		**JVT-O016**	33.45	37.69	38.39	10.07%	32.04
	Compare	0.05	-0.05	-0.01	1.08%	-0.01		***Compare***	0.13	-0.06	-0.08	-2.91%	0
Paris	**Proposed**	**31.37**	**34.63**	**34.94**	**4.87%**	**64.03**	**Suzie**	**Proposed**	**31.18**	**40.28**	**39.94**	**18.18%**	**12.08**
	JVT-G012	30.83	34.46	34.74	56.65%	64.84		**JVT-G012**	30.95	40.2	39.79	104.37%	12.13
	Gains	0.54	0.17	0.2	-51.78%	-0.81		***Gains***	0.23	0.08	0.15	-86.19%	-0.05
	JVT-O016	31.38	34.65	34.89	7.36%	64.16		**JVT-O016**	31.25	40.36	40.03	16.71%	12.06
	Compare	-0.01	-0.02	0.05	-2.49%	-0.13		***Compare***	-0.07	-0.08	-0.09	1.47%	0.02

0-7695-2650-0/06/$20.00 (c) IEEE

Authorized licensed use limited to: INSTITUTE OF COMPUTING TECHNOLOGY CAS. Downloaded on January 13, 2009 at 01:52 from IEEE Xplore. Restrictions apply.