NEW H.264 INTRA-RATE ESTIMATION AND INTER-RATE CONTROL DRIVEN BY IMPROVED MAD-BASED CONTRAST SENSITIVITY

M. Hrarti1, 4, A. Saadane1, 2, M. C. Larabi1, A. Tamtaoui3, D. Aboutajdine4

1Institut XLim, département Signal, Image et Communication, Université de Poitiers, UMR CNRS 6172
2Ecole Polytechnique de l’Université de Nantes, UMR CNRS 6597.
3LTI, INPT, 2, Avenue Allal El Fassi- Madinat Al Irfane- Rabat.
4GSCM-LRIT, B. P 1014, Faculté des Sciences de Rabat.

ABSTRACT

This paper aims to improve H.264 bit-rate control. The proposed algorithm is based on a new and efficient Rate-Quantization (R-Q) model for the intra frame. For the inter frame, we propose to replace the current use of MAD by a new MAD-based human Contrast Sensitivity (MAD-CS) which is a more accurate complexity measure. R-Q model for the intra frame results from extensive experiments. The optimal initial quantization parameter QP is based on both target bit-rate and complexity of I-frame. The I-frame target bit-rate is derived from the global target bit-rate by using a new non linear model. MAD-CS includes the contrast sensitivity of the human visual system and weights the absolute differences by the probability of their occurrence. Extensive simulation results show that the use of MAD-CS and the proposed R-Q model achieves better rate control for intra frames, reduces the bit-rates when compared to the H.264 rate control adopted in JM reference software, minimizes the peak to signal ratio variations among encoded pictures and increases significantly as well subjective visual quality (measured by psycho visual experiments) as objective one.

Index Terms— Bit allocation, H.264/AVC, Rate-Quantization model, MAD, Contrast Sensitivity Function.

1. INTRODUCTION

H.264/AVC standard improves significantly coding efficiency [1] and enables video to be compressed with much higher bit-rate reduction when compared to the previous standards such as MPEG2 and MPEG4. Rate control plays an important role in H.264 and has been the focus of several studies [2] [3] [4] [5]. Currently, three well-known rate control algorithms, MPEG2 TM5 [6], H.263 TMN8 [7] and MPEG4 VM8 [8], are widely used in practical applications. The rate control scheme used in H.264 remains more difficult than these latter since the quantization parameter (QP) has to be determined before the rate distortion optimization (RDO) is performed. To solve this “chicken and egg” dilemma, Li et al. [4] proposed a linear model for distortion prediction based upon complexity of prior frames. The resulting rate control algorithm has been adopted as an informative part by JVT for H.264 JM 10.2. Since, several studies [9] [10] [11] [12] have been conducted to improve the R-D performance of this rate control scheme and to solve its main defects. This paper deals with two of these defects. The first concerns the I-frames and the initial QP estimation. In Li’s scheme, the initial quantization step is selected from 10, 20, 25 and 35 according to target bit-rate and length of GOP. In [12], it is clearly shown that the initial QP is highly concerned with the sequence itself. Hence, without considering the complexity of I-frames, inappropriate initial QP can be selected and the errors caused by such a selection are propagated in the remaining frames, resulting in a low rate control performance and a poor visual quality. The second defect is the unsuitable bit allocation induced by the use of the MAD as complexity measure. Inaccurate MAD results in wrong QP and degrades both RDO performance and visual quality. In this paper, new approaches are proposed for these defects. First, the initial QP is determined by a mean of a new R-Q model derived from extensive experiments. In the proposed model, the optimal QP depends on both the target bit-rate and the complexity of the I-frame. Simulation results show that the bit-rate of the I-frame is better controlled and the quality of the following P frames of the GOP is improved. Secondly, we propose to replace the current MAD estimation by a new MAD-based human Contrast Sensitivity (MAD-CS). Simulation results show that with a low added computational complexity, MAD-CS correlates better with optimally allocated bits than the MAD.

The remainder of this paper is organized as follows. Section 2 describes the new R-Q model for the I-frame. Section 3 presents the proposed MAD-CS calculation. The experimental results are given in Section 4 and we conclude in Section 5.

2. PROPOSED I-FRAME R-Q MODEL

The optimal initial QP selection has been considered by several authors. Among them, two studies at least have properly tackled the problem. Pan et al. [16] and Zhou et al.
[12] have proposed methods where the initial QP takes into account target bit-rate, GOP length and an I-frame complexity measure. However, this complexity measure has been defined in the DCT domain in the first study whereas for the second it has been defined in the spatial domain but needs to buffer two frames at the beginning. In this section we develop a new R-Q model for I-frame.

In the first step of the proposed algorithm, a relationship between the I-frame target bit (RI) and the fixed global target bit (RT) is derived from a lot of experiments performed on three test sequences: Carphone, Foreman and Salesman. Each test video sequence is encoded at several bit-rates ranging from 10Kbps to 300Kbps. The actual bit-rates and two complexity measures (average gradient per pixel and Euclidian distance) of I-frame are computed. Then, the ratio between actual bit-rate and complexity measure of I-frame is plotted vs. global target bit. Whatever the complexity measure, the model which best fit the experimental data is the non linear model given by:

\[
\frac{RI}{\delta} = a \cdot \left( \frac{RT}{Fr} \right)^b
\]

(1)

Where \( \delta \) represents the I-frame complexity, \( Fr \) is the frame rate which equals to 30fps. The model parameters \( a \) and \( b \) are determined by the least squares method and are respectively equal to 123.4 and 0.38. Five other sequences have served to confirm the accuracy of the above relationship (Akiyo, Bridge-Close, Claire, Grandma and Miss-America).

Once the ratio \( RI/\delta \) is determined by the first model, we use a second model R-Q allowing the determination of the quantization parameter (QP) of the I-frame without having to calculate its complexity \( \delta \). To generate a relationship between the ratio \( RI/\delta \) and QP, we have used the same test conditions described above with an initial QP varying from 10 to 40 with a constant step of 2. Fitting the experimental data \( RI/\delta \) vs. QP leads to the non linear model of Eq.2. This model allows an optimal determination of I-frame initial QP without having to calculate its complexity unlike [12] and [16]:

\[
\frac{RI}{\delta} = \frac{1}{p_1 \cdot QP^2 + p_2 \cdot QP + p_3}
\]

(2)

\( p_1, p_2 \) and \( p_3 \) are the coefficients of R-Q model which are respectively equal to \( 2.443 \times 10^6, -6.934 \times 10^3 \) and \( 6.68 \times 10^3 \).

3. PROPOSED COMPLEXITY MEASURE

To represent the coding complexity, MAD is usually introduced for distortion measure instead of the variance of the coding unit (Frame or Macroblock (MB)). One way to minimize quality variation is to use more appropriate quality estimators than MAD. To solve this issue and balance the tradeoff between the algorithm complexity and the Rate-Distortion performance, we introduce three improvements in MAD calculation. Firstly, the absolute difference for each pixel in the MB (MAD) is computed based on the reconstructed image rather than on the predicted one. As the human visual system is sensitive to contrast changes rather than luminance ones, we calculate a local contrast \( C_{i,j} \) for each absolute difference pixel position \((i,j)\) in a macroblock.

Secondly, the sensitivity variation at different spatial frequencies is included by weighting \( C_{i,j} \) by the Contrast Sensitivity Function (CSF) to produce weighted contrast (C-Wi,j). The used CSF is isotropic and achromatic (luminance CSF) [13]. Thirdly, a spatial absolute difference pooling is performed to produce the new macroblock MAD measure. This is done by calculating the occurrence probability of difference contrasts weighted by CSF \( (PO(C-W_{i,j})) \). Finally, we use the Minkowski summation weighted by this occurrence probability as follow:

\[
MAD-CS = \left[ \frac{1}{256} \cdot \sum_{i=1}^{16} \sum_{j=1}^{16} (C-W(i,j) \cdot PO(C-W(i,j))) \right]
\]

(3)

4. EXPERIMENTAL RESULTS

The quality assessment has been performed at the macroblock level and the JVT reference software JM 10.2 [14] baseline profile has served for the different tests. The proposed approaches have been implemented using the same software, so that all parts of the encoder except for rate control are the same. Fourteen common test sequences have been used (Akiyo, Bridge-Close, Carphone, Claire, Coastguard, Container, Foreman, Grandma, Highway, Miss-America, Mother-Daughter, Salesman, Silent and Suzie) in QCIF format and 30 frames per second. Six bit-rates have been considered ranging from 10 to 80Kbps. We encode the first frame as I and the subsequent frames as P. We evaluate the experimental results with both subjective and objective assessment.

4.1. Subjective quality evaluation

For the subjective quality evaluation, we used the Double Stimulus Impairment Scale (DSIS) test method described in ITU-R BT.500-11 [15]. The scoring is done on discrete scale composed of 5 quality levels: excellent, good, fair, poor and bad that correspond respectively to scores of 5, 4, 3, 2 and 1. Fifteen observers were involved in the experiments. Table 1 lists the evaluation results only when MAD-CS estimation is implemented. Higher the score is, better is the obtained perceptual quality. From Table 1, we can notice that the proposed approach is visually efficient since it implies at least a score gain of 0.36 in comparison to the JM 10.2.
Table 1: Subjective scores using MAD-CS estimation

<table>
<thead>
<tr>
<th>Sequence</th>
<th>DSIS Score JM10.2</th>
<th>MAD-CS Score</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo (30Kbps)</td>
<td>3.57</td>
<td>4.36</td>
<td>+0.79</td>
</tr>
<tr>
<td>Bridge Close (80Kbps)</td>
<td>4</td>
<td>4.36</td>
<td>+0.36</td>
</tr>
<tr>
<td>Carphone (80Kbps)</td>
<td>4.07</td>
<td>4.64</td>
<td>+0.57</td>
</tr>
<tr>
<td>Mother Daughter (30Kbps)</td>
<td>2.57</td>
<td>2.93</td>
<td>+0.36</td>
</tr>
<tr>
<td>Silent (45Kbps)</td>
<td>2.64</td>
<td>3.29</td>
<td>+0.65</td>
</tr>
</tbody>
</table>

4.2. Objective quality evaluation

To evaluate the overall performance of our proposed approaches for all test sequences, we use the average PSNR and the average bit-rate difference. These measures are often used in the compression literature to compare Rate-Distortion (RD) performance between two different methods. The results are shown in Table 2 including those related to the MAD-CS alone and the combination with the proposed I-Frame R-Q Model. This allows noticing that for both proposed approaches, the obtained results are better than the reference software JM10.2 in terms of PSNR gain (up to 0.65dB for MAD-CS and up to 2.09dB for the combined approach). Of course, the PSNR increase leads to a visual quality improvement but another aspect needs to be highlighted. It concerns the gain of bit-rate that has been achieved for the major part of the test sequences. The amount of this gain varies reaches 11.3 Kbps (25.1%) for Claire at a target bit-rate of 45Kbps.

Figure 1 gives the rate distortion curves (RD) for both Grandma (a) and Bridge-Close (b) sequences when our combined strategy is applied. It is shown that the proposed approach achieves a higher PSNR for all selected bit-rates ranging from 10 to 80Kbps in comparison with the JM10.2 rate control.

Figure 2 depicts the variation of the PSNR for all the frames of Akiyo and by comparing our combined strategy to the JM10.2. We can remark that the proposed method produces a significant improvement in PSNR (Average PSNR gain of 0.84dB).

Finally, Figure 3 represents the buffer level of Akiyo sequence encoded at 80Kbps and using 3 GOPs when the combined strategy is applied in comparison to the JM10.2 rate control. On one hand, it can be shown that the buffer level when our approaches are applied is higher than JM10.2 buffer level at the beginning of the first GOP. This is due to the fact that the initial I-frame and a few P-frames are allocated with more bits than the average, which can increase the R-D performance of the whole sequence. On other hand, our combined strategy maintains a less and steadier buffer level than that for JM10.2.

5. CONCLUDING REMARKS

In this paper, we have proposed two approaches that aim to correct two defects in H.264 bit-rate control (JM 10.2): the inappropriate initial QP for I-frame and the inaccurate complexity estimation (MAD). The first approach proposed a new and efficient Rate-Quantization (R-Q) model that determines an optimal initial QP according to both target bit-rate and complexity of the I-frame without having to determine this latter. The second approach intends to replace the current inter coding unit complexity (MAD) by a new MAD-based Human Contrast Sensitivity (MAD-CS) which allows more accurate bit allocation. Objective and subjective experiments show that the proposed approaches improved significantly the R-D performance in low bit-rate, reduced PSNR and bit-rate variations among frames and obviously improved the perceptual quality. Therefore, this approach could open new directions as one can expect more performances if more HVS characteristics are considered.

Figure 1: Average PSNR results for sequences (a) “Grandma” and (b) “Bridge-close” encoded at various bit-rates when MAD-CS estimation and the new I-Frame Model R-Q are applied.
6. REFERENCES


Table 2: Comparison of Average PSNR and Bit-Rate Differences in both: MAD-CS and MAD-CS combined with R-Q Model.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Target Bit-rate (Kbps)</th>
<th>JM10.2 Average PSNR (dB)</th>
<th>JM10.2 Average</th>
<th>Gain</th>
<th>JM10.2 Average</th>
<th>Gain</th>
<th>JM10.2 Average</th>
<th>Gain</th>
<th>JM10.2 Average</th>
<th>Gain</th>
<th>JM10.2 STD of PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo</td>
<td>80</td>
<td>42.37</td>
<td>42.98</td>
<td>+0.61</td>
<td>43.21</td>
<td>+0.84</td>
<td>80.41</td>
<td>2.93</td>
<td>79.39</td>
<td>0.73</td>
<td>0.71</td>
</tr>
<tr>
<td>Bridge</td>
<td>45</td>
<td>32.36</td>
<td>32.99</td>
<td>+0.13</td>
<td>34.42</td>
<td>+1.56</td>
<td>45.38</td>
<td>0.07</td>
<td>45.45</td>
<td>-0.07</td>
<td>0.64</td>
</tr>
<tr>
<td>Close</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claire</td>
<td>30</td>
<td>38.72</td>
<td>39.37</td>
<td>+0.65</td>
<td>39.97</td>
<td>+1.25</td>
<td>30.50</td>
<td>0.15</td>
<td>30.37</td>
<td>0.13</td>
<td>0.95</td>
</tr>
<tr>
<td>45</td>
<td>40.6</td>
<td>41.03</td>
<td>40.6</td>
<td>0</td>
<td></td>
<td></td>
<td>45.81</td>
<td>0.16</td>
<td>45.65</td>
<td>1.14</td>
<td>1.4</td>
</tr>
<tr>
<td>Grandma</td>
<td>45</td>
<td>35.97</td>
<td>36.34</td>
<td>+0.37</td>
<td>38.06</td>
<td>+2.09</td>
<td>45.22</td>
<td>0.03</td>
<td>45.29</td>
<td>0.07</td>
<td>0.94</td>
</tr>
<tr>
<td>45</td>
<td>39.79</td>
<td>40.29</td>
<td>40.50</td>
<td>+0.5</td>
<td>40.71</td>
<td>+0.71</td>
<td>80.30</td>
<td>0.25</td>
<td>80.16</td>
<td>0.14</td>
<td>1.12</td>
</tr>
<tr>
<td>Miss America</td>
<td>30</td>
<td>39.19</td>
<td>39.44</td>
<td>+0.25</td>
<td>39.95</td>
<td>+0.76</td>
<td>30.33</td>
<td>0.16</td>
<td>30.17</td>
<td>0.17</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Figure 2 – PSNR for Akiyo sequence encoded at 80 Kbps and 30 fps when MAD-CS estimation and the new I-Frame Model R-Q are applied.

Figure 3 – Buffer Level of Akiyo sequence encoded at 80Kbps, 30 fps and using 3 GOPs when MAD-CS estimation and I-Frame Model R-Q are applied.