MODEL BASED ANALYSIS FOR QUANTIZATION PARAMETER CASCADING IN HIERARCHICAL VIDEO CODING

Xiang Li[12], *Peter Amon*[2], *Andreas Hutter*[2], *and André Kaup*[1]

[1]Chair of Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg, Erlangen, Germany
[2]Siemens Corporate Technology, Information & Communications, Munich, Germany

ABSTRACT

Originally, the hierarchical coding structure was proposed to achieve temporal scalability. Soon after, it was realized that with a proper quantization parameter cascading (QPC) scheme the general performance can be significantly improved by hierarchical coding. However, the theory behind the gain has not been explored so far. In this paper, the QPC in hierarchical coding is investigated by model based emulations. From the analysis, it is noticed that a parameter β which represents the error propagation in a group of pictures greatly affects the performance of QPC in hierarchical coding. Based on β, a simple adaptive QPC algorithm is designed. Simulations verify the efficiency of this algorithm: a gain up to 0.5 dB is obtained over the most recent SVC reference software.

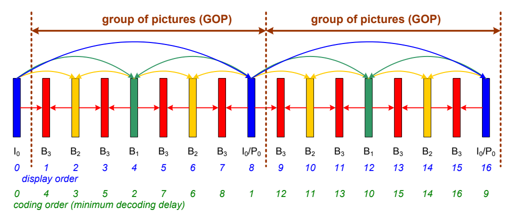
Index Terms— Video coding, Optimization methods

1. INTRODUCTION

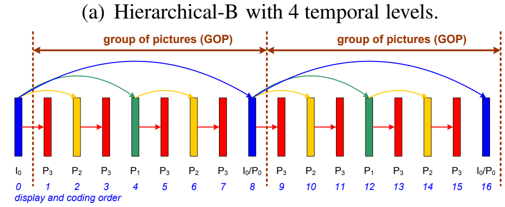
In recent years, scalable video coding (SVC) has been a hot research topic. During the development of the SVC [1], motion compensated temporal filtering (MCTF) [2] was first proposed to achieve temporal scalability. However, MCTF is a quite complex coding structure. More important, since the open-loop coder control of an MCTF encoder cannot compensate for quantization errors of the reference pictures, MCTF does not improve the coding efficiency much [3]. Therefore, MCTF was replaced by a simpler scheme, i.e., the hierarchical-B (H-B) coding structure.

A typical (H-B) coding structure with 4 dyadic levels is depicted in Fig. 1(a). Basically, coding with the H-B structure is the same as the traditional IBBP coding except that in motion compensation a frame in a certain temporal level cannot use frames from higher levels as references [3]. Although such a kind of restriction may degrade the overall coding efficiency, it guarantees the temporal scalability since frames in higher levels can be discarded without impact on those in lower levels.

Nevertheless, to achieve a certain temporal scalability, such as 4 temporal levels, a relatively big GOP size has to be employed, which leads to a large delay since the I/P frame



(a) Hierarchical-B with 4 temporal levels.



(b) Hierarchical-P with 4 temporal levels.

Fig. 1. Hierarchical Coding Structure

in the next GOP must be coded before the B frames in the current GOP. Accordingly, the hierarchical-P (H-P) coding structure was proposed. As shown in Fig. 1(b), by replacing the B frames with P frames, H-P keeps the temporal scalability while avoiding the delay. Therefore, it is quite appealing to real-time communication applications.

More important, it was observed that with a proper QP (quantization parameter) cascading (QPC) scheme, the hierarchical coding structures (including H-B and H-P) can significantly improve the overall coding efficiency [3]. For example, the QPC scheme in the current SVC reference software is as follows [3, 4],

$$QP_k = \begin{cases} QP_{k-1} + 4, & k = 1 \\ QP_{k-1} + 1, & k > 1 \end{cases} \tag{1}$$

where k denotes the temporal level. Although such a kind of scheme is efficient, the theory behind the gain has not been explored so far. Intuitively, a better understanding on this problem will lead to a better QPC scheme which will further improve the coding efficiency. Therefore in this paper, the performance of QPC scheme is analyzed, based on which an adaptive QPC algorithm is proposed. Simulations show that a gain up to 0.5 dB over the SVC reference software can be obtained.

The rest of this paper is organized as follows. First, the behavior of QPC in hierarchical coding is investigated by model based emulations in Section 2. From the analysis, it is observed that a parameter β which represents the error propagation in a GOP greatly affects the performance of QPC. Based on the value of β, a simple adaptive QPC algorithm is developed in Section 3. After that, the performance of the adaptive QPC is verified by simulations in Section 4. Finally, the conclusions and future work are presented in Section 5.

2. MODEL BASED ANALYSIS FOR QP CASCADING

Intuitively, the performance of QPC in hierarchical coding should be related to the correlation among frames. However, how to measure this correlation is unknown. In this section, the performance of QPC is first formulated in R-D (Rate-Distortion) sense. Then its behavior is investigated by model based emulations from which a parameter representing the frame correlation is discovered.

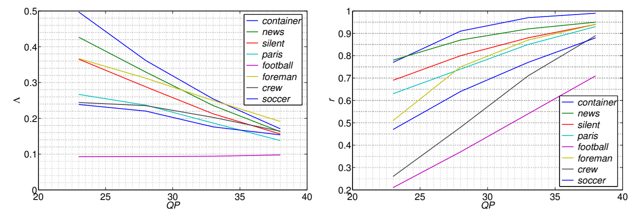
2.1. Problem Formulation

Theoretically, a large PSNR fluctuation will occur when QP varies much in successive frames, such as the scheme proposed in (1). However, [3] claimed that the reconstructed video was still subjectively smooth. Considering there is no widely accepted subjective quality measurement so far, the following discussion on coding efficiency is always in a R-D sense. The visual quality is assumed to be in line with the related PSNR.

In our previous work on frame level rate-distortion optimization [5], accurate R-D models were developed based on the Laplace distribution of transformed residues. For convenience, the closed form of the models are listed as follows,

$$\begin{aligned}
f_R(\Lambda, Q, r) = \frac{Se^{-\xi\Lambda Q}}{\ln 2}\{&(1-e^{-(1-\gamma)\Lambda Q})[\ln(1-e^{-(1-\gamma)\Lambda Q}) \\
&\cdot r - (1-r)\ln(1-r)] - (1-e^{-(1-\gamma)\Lambda Q}) \\
&\cdot \ln(1-e^{-(1-\gamma)\Lambda Q}) + [1-r(1-e^{-(1-\gamma)\Lambda Q})] \\
&\cdot \ln[1-r(1-e^{-(1-\gamma)\Lambda Q})] + e^{-(1-\gamma)\Lambda Q}(\ln 2 \\
&- \ln(1-e^{-\Lambda Q}) - \gamma\Lambda Q + \frac{\Lambda Q}{1-e^{-\Lambda Q}})\}, \\
f_D(\Lambda, Q) = \frac{\Lambda Q \cdot e^{\gamma\Lambda Q}(2+\Lambda Q-2\gamma\Lambda Q)+2-2e^{\Lambda Q}}{\Lambda^2(1-e^{\Lambda Q})}&,
\end{aligned} \tag{2}$$

where $f_R(\cdot)$ and $f_D(\cdot)$ represent the average rate and distortion for a single frame, S and ξ are two constants, γ indicates the ratio between the rounding offset and the quantization step, Λ represents the energy of transformed residues ($\Lambda = \sqrt{2}/\sigma$ and σ is the standard deviation of the transformed residues), Q is the quantization step which is defined in (3) according to [4, 6], and r denotes the ratio between the percentage of skipped MBs and that of the quantized zeros in a single frame. It should be noted that in JSVM [4], $\gamma = 1/3$ for temporal level 0 and $\gamma = 1/6$ for others. In addition, $0 \le r \le 1$ since the percentage of skipped MB (the residues



(a) Λ vs QP. (b) r vs QP.

Fig. 2. Relationship between QP, Λ and r (Eight sequences from [7, 8], Hierarchical-P structure, I+128P frames, GOP=8, no QP cascading, $QP = 23, 28, 33, 38$, each point represents the average of 128 P frames).

in the MB are all quantized zeros) is always no more than that of quantized zeros.

$$Q = 2.5 \cdot 2^{QP/6-2}. \tag{3}$$

Based on these R-D models, the average rate and distortion for frames in a GOP can be described as follows,

$$R = (\sum_{k=0}^{L} n_k \cdot f_R(\Lambda_k, Q_k, r_k)) / \sum_{k=0}^{L} n_k, \tag{4}$$

$$D = (\sum_{k=0}^{L} n_k \cdot f_D(\Lambda_k, Q_k)) / \sum_{k=0}^{L} n_k, \tag{5}$$

where L is the number of temporal levels, n_k denotes the number of frames in the temporal level k, and Λ_k, Q_k and r_k are the related parameters for temporal level k.

Putting (3) into (4) and (5), the optimization for a GOP is formulated as finding a set of $\{QP_k\}$ for the GOP so that the overall R-D performance is optimal. However, to mathematically solve such a kind problem is almost impossible since it not only is a constrained problem, but also involves many control parameters, such as Λ_k, Q_k and r_k for each temporal level. Therefore in this paper, the effects of all parameters are investigated by model based emulations, which is discussed in the next subsection.

2.2. Model Based R-D Curve Emulation

To evaluate the overall R-D performance, the most widely accepted approach is comparing the related R-D curves. Therefore in this subsection, the effects of parameters in QPC scheme are investigated by emulating R-D curves.

According to the suggestions by the video coding expert group (VCEG) of ITU-T [7], average PSNR is employed as the distortion measure in this paper. Thus D_{PSNR} is used instead of (5) in the following discussions, where

$$D_{\text{PSNR}} = (\sum_{k=0}^{L} n_k \cdot 10\log\frac{255^2}{D(\Lambda_k, Q_k)}) / \sum_{k=0}^{L} n_k. \tag{6}$$

To emulate R-D curves, Λ_k and r_k for each temporal level have to be determined. Simulations (as shown in Fig. 2) indicate that these two parameters can be approximated by piecewise linear functions of QP, i.e.,

$$\begin{aligned}
\Lambda &= \Lambda_a + \alpha_\Lambda \cdot (QP - QP_a), \\
r &= r_a + \alpha_r \cdot (QP - QP_a),
\end{aligned} \tag{7}$$

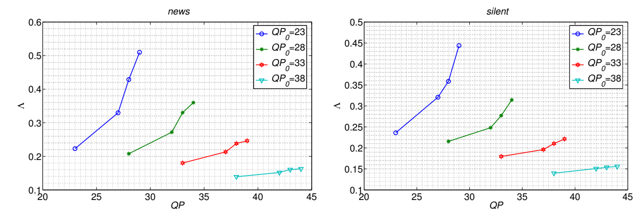


Fig. 3. Relationship between QP, Λ_0 and Λ_k. (GOP=8, 4 temporal levels, $QP_0 = 23, 28, 33, 38$, each point represents the average of P frames in the same temporal level.)

where Λ_a and r_a denote the related Λ and r at anchor QP_a, α_Λ and α_r represent the slopes of the corresponding curves.

Intuitively, Λ_k and r_k are highly related to Λ_0 and r_0. Although both of them can be approximated by (7), Λ_k is statistically better to be estimated by

$$\Lambda_k = \Lambda_0 + \beta \cdot (QP_k - QP_0), \tag{8}$$

where β is a linear factor which represents the error propagation introduced by the quantizations within the GOP. In fact, this approximation can be observed from Fig. 3, where the 4 points in each curve corresponds to the 4 temporal levels in one QPC scheme. Basically, a positive β indicates that almost no error propagation is caused by large quantizations in this GOP since the standard deviations of transformed residues in the frames of this GOP become smaller and smaller. On the contrary, a negative β denotes that the error propagation introduced by quantizations is serious, which may degrade the performance of motion compensation and the overall coding efficiency.

Based on (8), Λ_k and r_k for the temporal level k can be approximated by

$$\begin{aligned} \Lambda_k &= \Lambda_a + \alpha_\Lambda \cdot (QP_0 - QP_a) + \beta \cdot (QP_k - QP_0), \\ r_k &= r_a + \alpha_r \cdot (QP_k - QP_a), \end{aligned} \tag{9}$$

Plugging (3) and (9) into (4) and (6), the analytical R-D models for QP cascading scheme can be finally derived. Consequently, the related R-D curves can be emulated.

Fig. 4(a) shows the model based R-D curve emulation for *news*, where all parameters such as α_Λ and α_r are from practical statistics, "QPO:1,2,3" denotes the QP offsets from temporal level 1-3 to the temporal level 0 are 1, 2 and 3, respectively. To discover the parameter to which the R-D performance of QPC is most sensitive, different parameters are adjusted in emulations. Finally, it is observed that β is a key parameter. Fig. 4(b) shows the emulation result for a dummy sequence where $\beta = -0.01$ (other parameters are the same as those for *news*). Comparing the two figures, it can be observed that the performance of different QPC scheme is highly dependent on the value of β. When β is positive, i.e., almost no error propagation occurs, bigger QP offsets will lead to a higher coding efficiency since large quantizations will lead to a slightly higher distortion but a much lower rate. Otherwise, smaller QP offsets are preferred when β decrease to negative since error propagation in such a case is so serious

Algorithm 1: Proposed Adaptive QPC for a frame

input : $QP_k^f, \bar{\beta}^{j-1}, \beta_1^j, ..., \beta_{k-1}^j, \Delta_G$
output: QP_k
begin
  $\Delta_F \leftarrow 0$
  if $k = 0$ **then** `/*` Δ_G `depends on` $\bar{\beta}^{j-1}$ `*/`
    if $\bar{\beta}^{j-1} > 0$ **then** $\Delta_G + +$
    else if $\bar{\beta}^{j-1} < -0.005$ **then** $\Delta_G - -$
    $\Delta_G \leftarrow \max(-3, \min(3, \Delta_G))$
  else `/*` Δ_F `depends on` $\beta_1^j, ..., \beta_{k-1}^j$ `*/`
    $\beta_{temp} \leftarrow \frac{1}{k-1} \sum_{i=1}^{k-1} \beta_i^j$
    if $\beta_{temp} > 0$ **then** $\Delta_F + +$
    else if $\beta_{temp} < -0.005$ **then** $\Delta_F - -$
  end
  $QP_k \leftarrow QP_k^f + \Delta_G + \Delta_F$
end

that the overall coding efficiency may be greatly degraded by large quantizations.

Generally, the above observation is in line with the intuition that the correlation within the GOP determines the efficiency of QPC since error propagation is highly related to the correlation. More important, β can serve as a indicator for the error propagation. In the next section, a simple but efficient QPC scheme is designed based on β.

3. ADAPTIVE QP CASCADING SCHEME

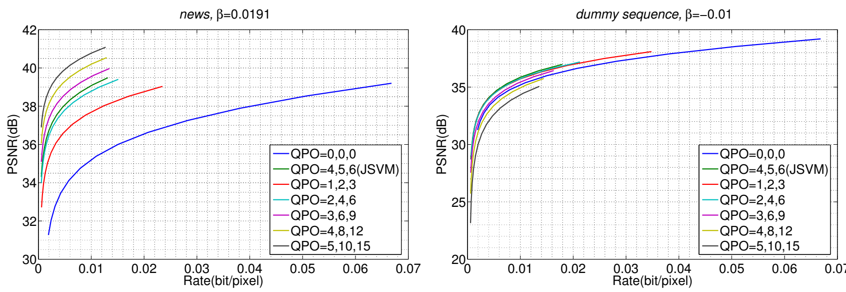
Intuitively, a more efficient QPC scheme can be achieved with the knowledge of the error propagation factor β. In this section, an adaptive QPC is designed based on a straightfoward application of β. It should be mentioned that this algorithm is to further prove the effectiveness of β. A more efficient QPC should be obtained by a better design.

The model based R-D curve emulation (Fig. 4) shows that when $\beta \geq 0$ a bigger QP offset is more efficient. Otherwise, a smaller offset should be preferred. Therefore, the proposed QPC is based on the value of β. To ease the implementation, β_k^j is defined as

$$\beta_k^j = (\bar{\Lambda}_k^j - \bar{\Lambda}_{k-1}^j)/(\overline{QP}_k^j - \overline{QP}_{k-1}^j), \tag{10}$$

where the superscript j indicates the GOP j, $\bar{\Lambda}_k^j$ and $\overline{QP}_k^j$ denote the average of available Λ and QP at temporal level k in the GOP j, respectively.

Specifically, for a frame at temporal level k in the GOP j, the proposed QPC is derived according to Algorithm 1, where QP_k^f is the fixed QPC scheme such as (1) in the JSVM software [4], Δ_G is a global offset which is initialized to 0 at the beginning of the sequence. From Algorithm 1, it can be noticed that the basic idea of the algorithm is increasing the QP offset when β is positive and decreasing it when β is below a certain threshold.



(a) (b)

Fig. 4. Model Based R-D Curve Emulation for *news* with 4 temporal levels. (According to practical statistics, $QP_a = 38, \Lambda_a = 0.1627, r_a = 0.95, \alpha_\Lambda = -0.017708, \alpha_r = 0.0112$)

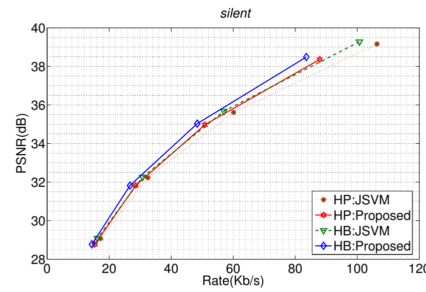


Fig. 5. Simulation result for *silent*.

Table 1. Average gain over JSVM 9.15

sequences	*news* (dB)	*silent* (dB)	*paris* (dB)	*city* (dB)	*crew* (dB)	*soccer* (dB)	**average** (dB)
H-B	0.28	0.32	0.21	-0.08	0.07	0.02	**0.14**
H-P	0.30	0.32	0.23	-0.07	0.03	0.02	**0.14**

4. SIMULATIONS AND DISCUSSIONS

The proposed adaptive QPC was verified by the most recent SVC referrence software JSVM 9.15 [4]. Six sequences (first three are from [7], others are from [8]) were tested. To enable a full hierarchical coding, 129 frames (I+16GOP, GOP=8, 4 temporal levels) of each sequences were coded with $QP_0 = 23, 28, 33, 38$. In addition, CABAC, fast search algorithm for motion estimation were enabled while quality scalability, spatial scalability, 8x8 transform, and low complexity MB mode were disabled.

Table 1 summaries the simulation results, which are the average gain in PSNR-Y over the JSVM software for each sequence. In fact, these gains measure the difference between two R-D curves [9]. On average, 0.14 dB gain over JSVM was obtained for both H-B and H-P. Particularly, the proposed algorithm is quite efficient for slow sequences, e.g., up to 0.5 dB gain can be observed for *silent* in Fig. 5. Moreover, the performance of "HP:Proposed" is similar to that of "HB:JSVM". Basically, this result is in line with the previous analysis. For slow sequence, the error propagation within a GOP is not sensitive to large quantizations (β is comparatively big) so that a significant gain can be achieved. While for fast/complex sequence such as *city* and *soccer* where the error propagation is quite serious for large quantizations, the proposed adaptive QPC shows a similar performance with the QPC scheme in JSVM.

5. CONCLUSIONS AND FUTURE WORK

In this paper, the overall performance of QP cascading in hierarchical coding structures H-P/B is investigated by model based R-D curve emulations. From the analysis, it is noticed that a parameter β which represents the error propagation within a GOP greatly affects the coding efficiency of QPC. Based on β, a simple QPC scheme is proposed. Simulations show that this algorithm is quite efficient for slow sequence and up to 0.5 dB gain was achieved, which further proves the effectiveness of the error propagation factor β.

For the next step, extending this model based analysis to a more complex scenario, such as from a single GOP to multiple GOPs or from single layer to multiple layers, is an interesting topic. We believe that the model based analysis will contribute much for the improvement of the overall coding efficiency.

6. ACKNOWLEDGMENT

This work was achieved with help of the European Community's Seventh Framework Program through grant agreement ICT OPTIMIX n°INFSO-ICT-214625.

7. REFERENCES

[1] G. J. Sullivan, T. Wiegand, and H. Schwarz, "ITU-T Rec. H.264 — ISO/IEC 14496-10 Advanced Video Coding Defect Report (JVT-Z210)," in *JVT Meeting (Joint Video Team of ISO/IEC MPEG & ITU-T VCEG)*, Antalya, Turkey, Jan. 2008.

[2] J.R. Ohm, "Advances in scalable video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 42–56, Jan. 2005.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF," in *IEEE Int. Conf. on Multimedia & Expo (ICME)*, Toronto, Canada, Jul. 2006, pp. 1929–1932.

[4] JVT, "H.264/SVC reference software (JSVM 9.15) and manual," CVS sever at garcon.ient.rwth-aachen.de, Sep. 2008.

[5] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, accepted in Jun., 2008.

[6] JVT, *Advanced Video Coding (AVC) - 3rd Edition*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10). 2004.

[7] T. Tan, G.J. Sullivan, and T. Wedi, "Recommended simulation common conditions for coding efficiency experiments revision 1 (VCEG-AE010)," in *ITU-T VCEG SG16 Q.6*, Jan. 2007.

[8] M. Wien and H. Schwarz, "Testing Conditions for SVC Coding Efficiency and JSVM Performance Evaluation (JVT-Q205)," in *JVT Meeting (Joint Video Team of ISO/IEC MPEG & ITU-T VCEG)*, 2005.

[9] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves (VCEG-M33)," in *VCEG Meeting (ITU-T SG16 Q.6)*, Austin, Texas, USA, Apr. 2001.