

Adaptive Quantization Parameter Cascading for Hierarchical Video Coding

Xiang Li^{*†}, Peter Amon[†], Andreas Hutter[†], and André Kaup^{*}

^{*}Chair of Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg, Erlangen, Germany

[†]Siemens Corporate Technology, Information & Communications, Munich, Germany

Abstract—Quantization parameter (QP) cascaded hierarchical prediction structures have been proved as efficient techniques in hybrid video coding. However, the current QP cascading method is empirical and not adaptive. The reason for the higher coding efficiency of this method has not been fully explored so far. In this paper, the rate-distortion performance of QP cascaded hierarchical video coding is first analyzed with dependent rate-distortion function. Then the optimal offset in linear QP cascading scheme is derived theoretically. It is shown that the widely accepted empirical QP cascading method is actually an approximation to the theoretical solution in fast movement environment. For slow sequences, an average gain of 0.43 dB can be achieved by the combination of two proposed adaptive algorithms.

I. INTRODUCTION

During the development of scalable video coding (SVC), quantization parameter (QP) cascaded hierarchical prediction structures (hierarchical-P and hierarchical-B) [1] were proposed. In principle, hierarchical coding is in the same manner as that of the traditional IPPP and IBBP schemes except that in motion compensation a frame in a certain temporal level cannot reference to the frames from higher levels [1]. With such a restriction, temporal scalability is guaranteed since the decoding of the frames in lower levels will not be impacted even if the frames in higher levels are discarded.

Due to the hierarchical prediction structures, frames at different temporal levels are with different importance in motion compensation process. Accordingly, it is natural to weight different temporal levels with different QPs during the encoding, which leads to a technology named QP cascading (QPC) [1]. Based on experimental summarization, an empirical QPC method for hierarchical prediction structures was proposed [1], i.e.,

$$\Delta QP_k = 4 + (k - 1), \quad k > 0, \quad (1)$$

where ΔQP_k denotes the QP offset from the temporal level k to the temporal level 0. It is shown that such QP cascaded hierarchical prediction structures (QPC-HPS) can significantly improve the overall coding efficiency. More important, it can be applied not only in SVC, but also in single layer video coding. Therefore, QPC-HPS draws much attention.

Basically, the empirical QPC in (1) can be generalized to a linear scheme, i.e.,

$$QP_k = \begin{cases} QP_0 & k = 0 \\ QP_0 + b + m \cdot (k - 1) & k > 0, \end{cases} \quad (2)$$

where QP_k represents the QP for the temporal level k , b is the QP offset base, and m represents the incremental QP

for temporal levels above 1. Simulations show that the offset b affects the overall performance greatly. Moreover, a fixed QP_0 in (2) may not be efficient in practice. Clearly, how to properly select b and QP_0 is a key problem in QPC-HPS. Therefore in this paper, two adaptive methods are developed. First, the optimal b is investigated with the dependent rate-distortion (R-D) function proposed in our previous work [2]. Then an adaptive QP_0 selection method is discussed based on our previous analysis on QPC-HPS [3].

The rest of this paper is organized as follows. First, the R-D models for linear QPC-HPS are developed in Section II. Then the two proposed adaptive methods are discussed in Section III and experimentally verified in Section IV. Finally, the conclusions and future work are presented in Section V.

II. R-D MODELS FOR LINEARLY QP CASCADED HIERARCHICAL PREDICTION STRUCTURES

To investigate the R-D performance of QPC-HPS, the inter-frame dependency has to be counted. Therefore in this section, our previously proposed dependent R-D function is first reviewed in Section II-A. Based on it, the R-D models for linear QPC-HPS are derived in Section II-B and Section II-C, respectively. Considering that GOP is a relatively stable and independent coding unit for hierarchical prediction structures, both R and D models are at GOP level.

A. Review of Dependent R-D Function

To consider the inter-frame dependency in budget allocation for a group of frames, a dependent R-D function was proposed in our previous work [2].

Taking S_n , the percentage of skipped 8x8 luminance blocks [4] in reconstructed frame $\hat{\mathbf{I}}_n$ as the quantitative measure for the inter-frame dependency between frame $\hat{\mathbf{I}}_n$ and its reference frame, the distortion D_n for $\hat{\mathbf{I}}_n$ can be calculated as

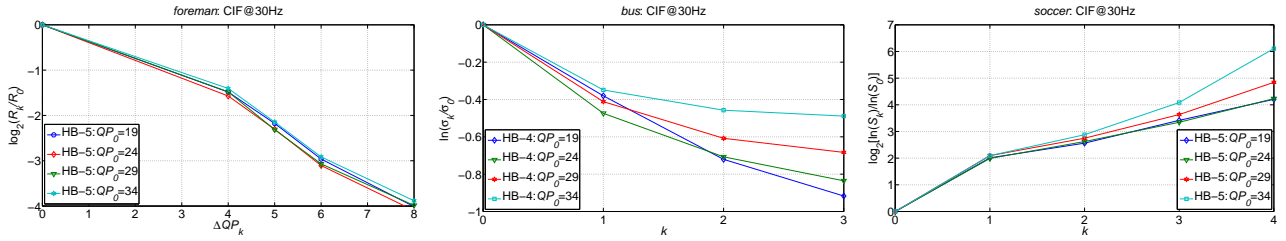
$$D_n = S_n \cdot D_n^S + (1 - S_n) \cdot D_n^N, \quad (3)$$

where D_n^S and D_n^N denote the average distortion for the skipped and non-skipped regions in $\hat{\mathbf{I}}_n$, respectively.

When $\hat{\mathbf{I}}_n$ is a P frame which is totally dependent on its reference frame $\hat{\mathbf{I}}_p$, the distortion D_n^{SP} can be derived as

$$\begin{aligned} D_n^{SP} &= f_{DS}(\hat{\mathbf{I}}_n) = (\hat{\mathbf{I}}_n - \mathbf{I}_n) \cdot (\hat{\mathbf{I}}_n - \mathbf{I}_n) \\ &= (\hat{\mathbf{I}}_p - \mathbf{I}_n) \cdot (\hat{\mathbf{I}}_p - \mathbf{I}_n) \approx (\hat{\mathbf{I}}_p - \mathbf{I}_p) \cdot (\hat{\mathbf{I}}_p - \mathbf{I}_p) = D_p, \end{aligned} \quad (4)$$

where $f_{DS}(\hat{\mathbf{I}}_n)$ represents the function for D_n^S , “ \cdot ” denotes the dot product, $\hat{\mathbf{I}}_n = \hat{\mathbf{I}}_p$ is due to the complete dependency,



(a) Relationship between (R_k/R_0) and ΔQP_k . (b) Relationship between (σ_k/σ_0) and k . (c) Relationship between (S_k/S_0) and k .

Fig. 1. R-D Approximations. (Hierarchical-B, 129 frames, I+16GOP or I+8GOP, each point represents the average of frames at the same temporal level.)

81 and the approximation is from the assumption that the source
82 frame \mathbf{I}_n and its reference \mathbf{I}_p are similar, finally D_p represents
83 the distortion of $\hat{\mathbf{I}}_p$.

84 In contrast, when $\hat{\mathbf{I}}_n$ is a B frame, the distortion D_n^{SB} can
85 be obtained by

$$\begin{aligned} D_n^{SB} &= f_{DS}(\hat{\mathbf{I}}_n) = [(\hat{\mathbf{I}}_p + \hat{\mathbf{I}}_b)/2 - \mathbf{I}_n] \cdot [(\hat{\mathbf{I}}_p + \hat{\mathbf{I}}_b)/2 - \mathbf{I}_n] \\ &= (\hat{\mathbf{I}}_p - \mathbf{I}_n) \cdot (\hat{\mathbf{I}}_p - \mathbf{I}_n)/4 + (\hat{\mathbf{I}}_b - \mathbf{I}_n) \cdot (\hat{\mathbf{I}}_b - \mathbf{I}_n)/4 \quad (5) \\ &\quad + (\hat{\mathbf{I}}_p - \mathbf{I}_n) \cdot (\hat{\mathbf{I}}_b - \mathbf{I}_n)/2 \approx (D_p + D_b)/4, \end{aligned}$$

86 where $\hat{\mathbf{I}}_p$ and $\hat{\mathbf{I}}_b$ represent the two references for $\hat{\mathbf{I}}_n$, D_p and
87 D_b are the related distortion, and the last term in the second
88 equation is regarded as zero since the two differential frames
89 $(\hat{\mathbf{I}}_p - \mathbf{I}_n)$ and $(\hat{\mathbf{I}}_b - \mathbf{I}_n)$ should be statistically uncorrelated.

90 Assuming that the non-skipped region of $\hat{\mathbf{I}}_n$ is a Gaussian
91 source, D_n^N is derived from the rate-distortion function of
92 Gaussian source [5], i.e.,

$$D_n^N = f_{DN}(\sigma'_n, R_n) = \sigma'^2_n \cdot 2^{-2 \cdot R_n}, \quad (6)$$

93 where $f_{DN}(\cdot)$ indicates the function for D_n^N , σ'_n denotes the
94 standard deviation of transformed residues in the non-skipped
95 region of $\hat{\mathbf{I}}_n$. Moreover, in (6) the budget for $\hat{\mathbf{I}}_n$ is totally spent
96 on the non-skipped region since the rate cost for skipped 8x8
97 luminance blocks is negligible [4].

98 Plugging (4) (or (5)) and (6) into (3), the distortion D_n for
99 frame $\hat{\mathbf{I}}_n$ is obtained as

$$D_n = S_n \cdot f_{DS}(\hat{\mathbf{I}}_n) + (1 - S_n) \cdot f_{DN}(\sigma'_n, R_n). \quad (7)$$

100 As shown in [2], the performance of this dependent R-D
101 function (7) is quite good. Therefore, it will be employed
102 to model the R-D behavior of QPC-HPS in the following
103 subsections.

B. Rate Model for a GOP

104 There are many rate models proposed in the literature, such
105 as the quadratic model [6], and the Laplace distribution based
106 model [7]. To provide desired accuracy, control parameters
107 in addition to quantization interval are normally employed to
108 describe the properties of input video sequences. However,
109 these complex models are not well fit for the rate modeling in
110 this paper since for online video coding, the predictability of
111 the control parameters is even more important than the model
112 accuracy [8]. Moreover, these models will lead to complex
113 functions which are quite difficult to solve.

114 Generally in H.264/AVC [4] coding with IPPP/IBBP, the
115 output bitrate will be roughly reduced by half when the QP
116 value is increased by 6 [9]. However in QPC-HPS, the impact
117 of ΔQP_k on the bitrate R_k for the temporal level k is much

118 bigger. On average, the rate ratio R_k/R_0 is roughly reduced
119 by half when ΔQP_k is increased by 2.5, such as for *foreman*
120 in Fig. 1(a), namely,
121

$$R_k = R_0 \cdot 2^{-\Delta QP_k/2.5}. \quad (8)$$

122 Supposing that the average frame bit rate for a GOP with
123 K temporal levels (GOP size is 2^{K-1}) is \bar{R} , R_0 can be derived

$$R_0 = \frac{\bar{R} \cdot 2^{K-1}}{1 + \sum_{k=1}^{K-1} 2^{k-1} \cdot 2^{-\Delta QP_k/2.5}}, \quad (9)$$

124 where 2^{k-1} indicates the number of frames at the temporal
125 level k . Accordingly, the rate for each temporal level can be
126 determined by plugging (2) and (9) into (8).

C. Distortion Model for a GOP

127 Let D_k denote the average distortion for frames at temporal
128 level k , the total distortion D_S for a GOP can be derived as
129

$$D_S = D_0 + \sum_{k=1}^{K-1} [2^{k-1} \cdot D_k]. \quad (10)$$

130 As shown in Fig. 2, frames at temporal level k are normally
131 predicted from level $k-1$ in hierarchical-P (HP) or from
132 level $k-1$ and $k-2$ in hierarchical-B (HB). Although this
133 observation is not always true, such as for the first P_3 and
134 B_3 frames in Fig. 2, the percentage of exceptions is relatively
135 low. Therefore, keeping this observation as an assumption in
136 the derivation of distortion model will not lose much accuracy.
137 More important, with this assumption, a unified distortion
138 expression can be achieved so that the deduction process will
139 be much simplified.

140 Under the above prediction assumption, the dependent part
141 $f_{DS}(\hat{\mathbf{I}}_k)$ of the frame distortion D_k in HP and HB can be
142 derived as

$$D_k^P = S_k \cdot D_{k-1} + (1 - S_k) \cdot f_{DN}(\sigma'_k, R_k), \quad (11)$$

$$D_k^B = S_k \cdot \frac{D_{k-1} + D_{k-2}}{2} + (1 - S_k) \cdot f_{DN}(\sigma'_k, R_k), \quad (12)$$

143 where S_k and σ'_k are the inter-frame dependency and the
144 standard deviation of transformed residues in non-skipped
145 regions for temporal level k .

146 To obtain models for σ_k and S_k , simulations were con-
147 ducted. Fig. 1(b) show that the relationship between $\ln(\sigma_k/\sigma_0)$
148 and k can be linearly approximated, i.e.,
149

$$\ln(\sigma_k/\sigma_0) = c \cdot k, \quad (13)$$

149 where c is a constant for a sequence coding. Consequently,

$$\sigma_k = \sigma_0 \cdot \beta^k, \quad (14)$$

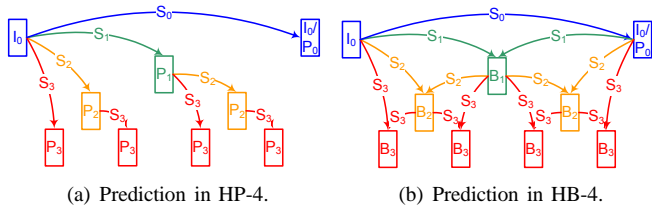


Fig. 2. Inter-frame dependency in hierarchical prediction structures

where β actually represents the error propagation between temporal levels in one GOP. Basically, $\beta < 1$ indicates a slight error propagation in the current GOP. In this case, bigger QP offsets are normally more efficient [3].

As to S_k , a more complex model is observed. Fig. 1(c) shows that an approximately linear relationship between $\log_2(\ln(S_k)/\ln(S_0))$ and k , namely

$$\log_2[\ln(S_k)/\ln(S_0)] = c' \cdot k, \quad (15)$$

where c' is a constant. Equivalently,

$$S_k = S_0^{2^{-\alpha \cdot k}}, \quad (16)$$

where α is a parameter indicating the increasing speed of S_k along the temporal level k . Typically, $\alpha = 1.5$ for HP/B-3/4/5.

Taking (8), (11) (or (12)), (14), and (16) into (10), the total distortion model D_S for a whole GOP can be determined.

III. ADAPTIVE QP CASCADING

In principle, the optimal offset b^* for linear QPC can be solved from the dependent rate-distortion function. However for QP_0 , it is difficult to directly model its effect on the overall R-D performance. Therefore, it is selected based on the error propagation within a GOP. In this section, two adaptive selection methods for these two parameters will be discussed in Section III-A and Section III-B, respectively.

A. R-D Optimal Offset in Linear QP Cascading

Based on the R-D function developed in the previous section, the R-D optimal offset b^* for linear QP cascading can be derived as the offset which minimizes the total distortion for the whole GOP under the constraint over the average rate.

Without loss of generality, assuming that the average frame bitrate is 1 bit/pixel, namely $\bar{R} = 1$, b^* can be calculated as

$$b^* = \arg \min_{R=1} \{D_S = D_0 + \sum_{k=1}^{K-1} [2^{k-1} \cdot D_k]\}. \quad (17)$$

Consequently, b^* can be determined by solving

$$\partial D_S / \partial b = 0. \quad (18)$$

As shown in the previous section, D_S is a complex model and so is the function (18). Although it is almost impossible to obtain a closed solution to (18), it can be numerically solved for given S_0 , α and β . Fig. 3 presents numerical solutions to HP/B-3/4/5 for several typical α and β values. For an easy understanding and implementation, the average inter-frame dependency in a GOP, namely S_G defined in (19), instead of S_0 is employed as the x-axis in the figures.

$$S_G = (S_0 + \sum_{k=1}^{K-1} 2^{k-1} \cdot S_0^{2^{-\alpha \cdot k}}) / 2^{K-1}. \quad (19)$$

From Fig. 3, it can be observed that the empirical method in (1) where $b^* = 4$ is actually an approximation to the theoretical solutions of HB/P-4/5. In fast movement environment such as those defined by JVT in [10], S_G is normally between 0.7 and 0.9 in the quality range of 30 – 40 dB. As shown in Fig. 3, when error propagation is slight ($\beta = 0.85$) $b \approx 4$ for such cases, which verifies that the proposed theory well matches the practice.

To ease the implementation of optimal offset b^* , lookup tables are employed. Considering HP/B-3/4/5 are widely used in practice, six lookup tables corresponding these six cases are built according to Fig. 3. For each table, 81 entries are evenly distributed to cover the range $S_G \in [0.15, 0.95]$. During the encoding, the first GOP is coded with the empirical QPC method in (1). From the second GOP, the proposed offset b^* is derived as follows. First, the S_G of the previous GOP is calculated as \bar{S}_p . Then the target index n to the related lookup table is obtained as

$$n = \min(15, \max(95, \lfloor 100 \cdot \bar{S}_p + 0.5 \rfloor)), \quad (20)$$

where $\lfloor y \rfloor$ denotes the floor function which keeps the integer part of y when $y \geq 0$. Finally, QP_k^* ($k > 0$) for the current GOP is derived as

$$QP_k^* = \max\{QP_k, \min\{QP_k + 6, QP_0 + T_{b^*}[n] + k - 1\}\}, \quad (21)$$

where QP_k denote the empirical method in (1), $T_{b^*}[n]$ indicates the lookup table, namely the values presented in Fig. 3.

It should be noted that although the derivation of b^* is complex, the proposed b^* selection method costs little computation and is easy to implement since it is based on lookup tables.

B. Adaptive Selection of QP_0

In addition to QP_k ($k > 0$), QP_0 for the temporal level 0 can be adaptive selected as well. Inspired by our previous work [3], the offset ΔQP_0 is adjusted according to the error propagation factor β in the current implementation, namely

$$\Delta QP_0^i \leftarrow \begin{cases} \min(\Delta QP_0^{i-1} + 1, 3) & \beta^{i-1} \leq 0.95 \\ \Delta QP_0^{i-1} & 0.95 < \beta^{i-1} \leq 1.1 \\ \max(0, \Delta QP_0^{i-1} - 1) & 1.1 < \beta^{i-1}, \end{cases} \quad (22)$$

where superscripts i and $i-1$ indicate the current and previous GOP, respectively. The refined QP_0^i for the current GOP is

$$QP_0^i = QP_0^f + \Delta QP_0^i, \quad (23)$$

where QP_0^f denotes the fixed QP_0 by the empirical method in (1) or (2).

Based on β , this adaptive QP_0 selection is with low computational complexity. Generally, the calculation of σ_k for each frame costs most computation in this method. When compared with motion estimation process, this computational payload is quite low.

IV. SIMULATIONS AND DISCUSSIONS

To verify the performance, the proposed AQPC algorithms were implemented in SVC reference software JSVM 9.15 [11]. HB/P-3/4/5 were tried for two set of sequences: Seq-AVC (*container, news, foreman*(QCIF), *silent, paris, tempete* defined in [12]) and Seq-SVC (*bus, football, foreman*(CIF), *mobile*,

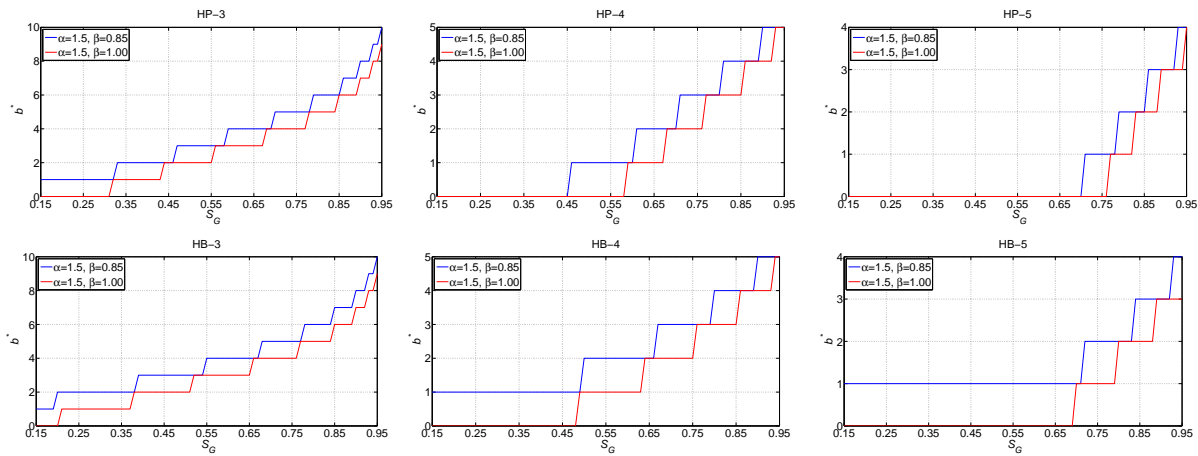


Fig. 3. Optimal offset b^* in linear QP cascading.

TABLE I

PERFORMANCE OF THE PROPOSED AQPC: AVERAGE GAIN (dB) FOR LUMINANCE COMPONENT IN BD-PSNR OVER JSVM 9.15

Sequence set	HB-3	HB-4	HB-5	HP-3	HP-4	HP-5	Average
Seq-AVC	0.41	0.39	0.16	0.43	0.39	0.14	0.32
Seq-SVC	0.01	0.00	-0.01	0.04	0.01	-0.03	0.00

232 *crew, city, harbour, soccer* defined in [10]). Totally 129 frames
233 of each sequences were coded with initial $QP_0 = 19, 24, 29, 34$.

234 Due to the limited space, only the average gains in BD-
235 PSNR [13] for the two set of sequences are presented in
236 Table I. On average, the combination of the two adaptive
237 methods AQPC obtains 0.32 dB for Seq-AVC while a similar
238 performance to the empirical method (1) for Seq-SVC. As
239 shown in the table, the gains are not evenly distributed:
240 More significant gains were achieved for slow sequences with
241 smaller GOP, such as 0.43 dB gain for HP-3 test on Seq-AVC.
242 In fact, such results are in line with the previous analysis.
243 For slow sequences, S_G is bigger than that in fast sequences.
244 Consequently, bigger QP offsets are employed and significant
245 gains are obtained. While for fast sequences and the case of
246 QPC with bigger GOP size, the QP offsets selected by the
247 proposed algorithms are close to those by the empirical method
248 (1), as shown in Fig. 3. Accordingly, similar performance were
249 observed.

V. CONCLUSIONS AND FUTURE WORKS

250
251 In this paper, two adaptive QP cascading algorithms for
252 hierarchical video coding are proposed. First, the R-D opti-
253 mal offset for linearly QP cascaded hierarchical prediction
254 structures is theoretically derived based on the dependent rate-
255 distortion function. It is shown that the empirical QPC method
256 employed in the SVC reference software is actually an practical
257 approximation to the theoretical scheme for the cases of
258 HB/P-4/5 in fast movement scenario. Second, an adaptive QP_0
259 selection is presented based on the error propagation within
260 a GOP. Comprehensive simulations show that the proposed
261 algorithms are more efficient than the reference software. For
262 slow sequences with smaller GOP size, over 0.4 dB gain on
263 average was obtained.

264 Nevertheless, the analysis in this paper is always in the
265 sense of rate-distortion. For the next step, how to count the

visual quality into the optimization process is an interesting
266 topic. Moreover, the optimal QPC scheme in the environments
267 of quality scalability and spatial scalability is also worth
268 investigation.
269

ACKNOWLEDGMENT

This work was achieved with help of the European Community's
270 Seventh Framework Program through grant agreement ICT OPTIMIX
271 n°INFSO-ICT-214625.
272
273

REFERENCES

- [1] 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. 275
- [2] X. Li, A. Hutter, and A. Kaup, "One-pass frame level budget allocation in video coding using inter-frame dependency," in *IEEE Int. Workshop on Multimedia Signal Process. (MMSP)*, Rio de Janeiro, Brazil, Oct. 2009, **accepted**. 276
- [3] X. Li, P. Amon, A. Hutter, and A. Kaup, "Model based analysis for quantization parameter cascading in hierarchical video coding," in *IEEE Int. Conf. on Img. Process. (ICIP)*, Cairo, Egypt, Nov. 2009, **accepted**. 277
- [4] *JVT, Advanced video coding (AVC) - 3rd edition*, ser. ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10), 2004. 278
- [5] T. Cover and J. Thomas, *Elements of Information Theory (Second Edition)*. Wiley-Interscience, 2006. 279
- [6] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997. 280
- [7] 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.*, vol. 19, no. 2, pp. 193–205, Feb. 2009. 281
- [8] J. Dong and N. Ling, "On model parameter estimation for H.264/AVC rate control," in *IEEE Int. Symposium on Circuits and Syst. (ISCAS)*, New Orleans, USA, May 2007, pp. 289–292. 282
- [9] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003. 283
- [10] 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)*, Poznan, Poland, Jul. 23-29 2005. 284
- [11] *JVT, "H.264 SVC reference software (JSVM 9.15) and manual"*, CVS sever at garcon.ient.rwth-aachen.de, Sep. 2008. 285
- [12] G. Sullivan, "Recommended simulation common conditions for H.26L coding efficiency experiments on low-resolution progressive-scan source material (VCEG-N81)," in *VCEG Meeting (ITU-T SG16 Q.6)*, Santa Barbara, USA, Sep. 24-27 2001. 286
- [13] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves (VCEG-M33)," in *VCEG Meeting (ITU-T SG16 Q.6)*, Austin, Texas, USA, Apr. 2-4 2001. 287