

Region-of-interest based rate control scheme for high efficiency video coding

Marwa Meddeb, Marco Cagnazzo, Béatrice Pesquet-Popescu

▶ **To cite this version:**

Marwa Meddeb, Marco Cagnazzo, Béatrice Pesquet-Popescu. Region-of-interest based rate control scheme for high efficiency video coding. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, May 2014, Florence, Italy. pp.7338 - 7342, 10.1109/ICASSP.2014.6855025 . hal-01088577

HAL Id: hal-01088577
https://hal.science/hal-01088577

Submitted on 1 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

REGION-OF-INTEREST BASED RATE CONTROL SCHEME FOR HIGH EFFICIENCY VIDEO CODING

Marwa Meddeb[1]*, Marco Cagnazzo*[1]*, and Béatrice Pesquet-Popescu*[1]

[1]Institut Mines-Télécom; Télécom ParisTech; CNRS LTCI

ABSTRACT

In this paper, we propose a new rate control scheme designed for the newest high efficiency video coding (HEVC) standard, and aimed at enhancing the quality of regions of interest (ROI). Our approach allocates a higher bit rate to the region of interest while keeping the global bit rate close to the assigned target value. This algorithm is developed for a videoconferencing system, where the ROIs (typically, faces) are automatically detected and each coding unit is classified in a region of the interest map. This map is given as input to the rate control algorithm and the bit allocation is made accordingly. Experimental results show that the proposed scheme achieves accurate target bit rates and provides an improvement in the region of interest quality, both in objective metrics and based on subjective quality evaluation.

Index Terms— HEVC, Rate control, ROI coding

1. INTRODUCTION

Rate control (RC) is an important tool that helps to deal with bit rate and compressed media quality fluctuations. RC methods have been widely studied and suitable schemes have been developed for specific applications [1]. This problem is also related to challenging issues such as resource availability, computational complexity and real-time [2]. In this paper, we consider RC for a specific class of applications, namely videoconferencing. In this context, one of the most interesting issues to focus on is the quality enhancement of regions of interest.

In videoconferencing systems (but also in other fields such as video surveillance and telemedicine), the subjective visual quality mainly depends on some important areas, called regions of interest (ROIs). Therefore, many contributions have introduced rate control algorithms based on ROIs. For example, in [3] a rate control scheme based on adjustable quality of the ROI has been proposed. The RC algorithm used the same quadratic model implemented in H.264/AVC to compute for each region a quantization parameter (QP) referring to a quality level chosen by the user. The same quadratic model is used in [4] to compute the QP of each macroblock and then adjust it referring to an input saliency map and to the number of bits allocated to each region. For a video surveillance system, RC in [5] uses a linear rate-quantization (R-Q) model to decide the bit-stream length and then the QP of each region.

These techniques considered the quadratic RC model and are useful for H.264/AVC implementations [3] [4]. Meanwhile, the new HEVC standard has been recently finalized by ITU-T and ISO/IEC [6] and many works have focused on rate control and developed new R-Q schemes for it. In the reference software two different schemes have been proposed. The first one is based on a quadratic model and the mean absolute difference (MAD) between the original and the reconstructed signal [7] [8]. In the second algorithm, an R-λ model that takes into account the hierarchical coding structure has been adopted [9]. Moreover, textured and non textured rate models for HEVC have been constructed to deal with more complex content and ensure more accurate rate control [10].

All the above-mentioned RC algorithms developed for HEVC do not take into account the importance of some regions of the frame. Therefore, we propose a new rate control scheme for videoconferencing systems at low bit rates which processes the faces and the background separately. Our proposed algorithm is based on the model implemented in the reference software HM.10 [9] and enhanced with three main features; first, using Viola and Jones object detection method [11], we detect our ROI and generate automatically an ROI map. The target bit rate is allocated for each region considering a fixed weight. Then, the QP of each coding unit (CU) is computed referring to the rate model of the corresponding region and the allocated bit budget. Finally, the proposed method considers independent rate-distortion models for each region and different clipping of QP variation, taking into account the importance of each part of the image. Overall, we show that the quality of the ROI is improved and the bit rate limit is respected.

The paper is organized as follows. Section 2 describes the general rate control problem and HEVC solutions. Then, the proposed rate control approach is explained in Section 3. Section 4 provides a description of experimental results of the proposed method. Finally, conclusions are given in Section 5.

2. RELATED WORKS

The objective of RC is to achieve a target bit rate as near as possible to a given constant with a minimum quality distortion.

Knowing that quantization consists in reducing the bit rate of the compressed video signal, the major role of RC algorithms is to find for each sample the appropriate QP under the constraint $R_s(\mathrm{QP}) \leq \mathrm{R_{max}}$. The fixed bit budget is $R_{\max}$ and $R_s(\mathrm{QP})$ is the number of coding bits for the source sample s.

In video coding, RC usually incorporates rate distortion optimization (RDO). Knowing the QP given by the rate control, RDO consists in minimizing the cost

$$J = D(\mathrm{QP}) + \lambda_{\mathrm{MODE}}\, R(\mathrm{QP}) \tag{1}$$

to achieve the optimized mode decision of each CU. Using a Lagrange multiplier λ_{MODE} in (1), the distortion $D(\mathrm{QP})$ is associated with the number of bits $R(\mathrm{QP})$ to evaluate all possible coding modes and select the one that minimizes J [12] [13].

Consequently, these problems need explicit rate-distortion models that relate the average bit rate to the QP. Several works have been done in perceptual quality and rate modeling. Different rate models have been developed, some of them based on simple linear expressions, others on more complex mathematical representations. For example, in [14], the traditional linear model that was employed in TM5 for high bit rate video coding is studied for HEVC; $R(\mathrm{QP}) = \frac{C}{\mathrm{QP}}$, where C is the model parameter. The quadratic model represented as

$$R(\mathrm{QP}) = \frac{C_1 \times \mathrm{MAD}}{\mathrm{QP}} + \frac{C_2 \times \mathrm{MAD}}{\mathrm{QP}^2} \tag{2}$$

where C_1 and C_2 are the model parameters, has been adopted in VM8 for MPEG4 [15], H.264/AVC [16] and also for HEVC [17].

The accuracy of these models has been enhanced by introducing the so-called complexity of the source, using the per pixel gradient value in the R-Q model in [18]. The sum of absolute transformed differences (SAD) has been adopted in [19]. The MAD between original frame and reconstructed one has been introduced in [7] as represented in (2). In a different way, the RC has improved by considering a representation in the ρ domain [20] as proposed in [21] and by taking into account additional parameters, as the frame rate [22].

The most recent rate-distortion model in the HEVC test software is the R-λ model expressed as follows:

$$\lambda = \alpha\, R^{\beta} \tag{3}$$

where α and β are the model parameters [9]. We note that this model defines a relationship between the rate in bits per pixel R and the Lagrange parameter λ which is used in RDO to decide the coding mode. Using this R-λ model, λ is generated first then the QP is computed. In our work, this model has been adopted and modified for our videoconferencing system.

For visual quality, a distortion model is usually developed to help predicting the relationship between the quality degradation and the quantization step. In fact, the model and the used metric vary from one work to another [1].

As stated before, each model targets a specific video coding system under particular conditions. However, all rate control methods aim at allocating appropriate number of bits and at determining the quantization parameter of each encoding unit. The complete R-λ RC scheme in HEVC can be represented as follows:

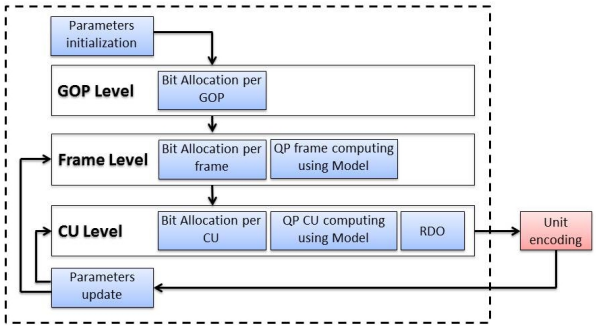


Fig. 1. Rate control scheme

As shown in Fig. 1, the controller operates at three main levels: group of pictures (GOP), frame and CU [6]:

i. GOP level: the input parameters are the target bit rate, the sequence frame rate, the GOP size and the virtual buffer occupancy. The rate control algorithm computes an average number of bits per GOP.

ii. Frame level: considering the average allocated bits per frame, a target bit is fixed for the current frame. The bit allocation takes into account the hierarchical level of each frame. Then, the R-λ model is used to compute the frame QP.

iii. CU level: the process is divided into three parts. First, the required number of allocated bits for the CU is computed using the frame budget, the cost of the coded CUs of the frame and the MAD of the CUs. Second, the budget is used in the R-λ model to compute λ and the QP of each CU. λ and QP variation are clipped referring to a fixed range. Finally, the last step is the RDO in order to find the optimized mode decision [23], referring to the QP given by the second step. The unit is then coded and all the parameters are updated.

3. PROPOSED APPROACH

The proposed approach is based on the R-λ model for HEVC. The relationship between R and λ represented by (3) in Sec. 2 is used to compute QP of the frame and each CU of the image. This model has given better performances than the quadratic one [7] [8]. In this section we describe the proposed approach in our work and we focus on the two main steps of the rate control: the bit allocation at frame and CU levels and the computation of QP by the proposed model.

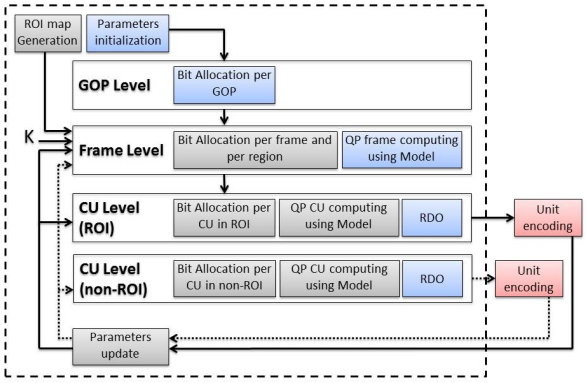


Fig. 2. ROI-based rate control scheme

Figure 2 shows the proposed ROI-based rate control scheme. The first step consists in detecting the faces in the scene and generating automatically an ROI map per frame, which will be given as input to our controller. The target bit rates allocated for the GOP and the current frame are obtained using the R-λ model. Then, the frame budget is divided into two parts according to a fixed factor K which is the ratio between the rate in bits per pixel of the ROI and the rate in bits per pixel of the rest of the frame (non-ROI). At the CU level, the ROI map is used to make a separate bit allocation for CUs of different regions. The R-λ model is applied for each CU using the allocated bit budget for the corresponding region (ROI or non-ROI). Once the CU is encoded, the model parameters of the corresponding region are updated, and the next CU is processed in a similar way.

3.1. Region bit allocation

We introduce the region bit allocation at two levels; at frame level to initialize a target amount of bits for each region, and at CU level to make independent bit allocation of CUs of different regions. At frame level, the positive constant K is selected. It represents the desired ratio between the ROI and non-ROI bit rates:

$$R_r = K \times R_n \tag{4}$$

where the subscript r denotes the ROI and n the non-ROI. We assume that the current allocated bits per frame T_{Pic} is the sum of the number of bits of the two regions T_r for the ROI and T_n for the non-ROI:

$$T_{\mathrm{Pic}} = T_r + T_n \tag{5}$$

$$T_n = R_n \times M \times P_n \tag{6}$$

where M is the total number of pixels of the frame and P_n the area of non-ROI. From (4), (5) and (6), non-ROI bit rate R_n is computed as follows:

$$R_n = \frac{T_{\mathrm{Pic}}}{M\ (1 + P_r\ (K-1))} \tag{7}$$

At CU level, the bit allocation depends on the number of bits allocated per region and on the weights of CUs of the same region. For each CU of the ROI the allocated bits are:

$$T_{\mathrm{CU}r} = \frac{T_r - T'_r}{\sum_{i \in I_r} w_{ir}}\ w_{\mathrm{CU}r} \tag{8}$$

where T'_r is the effective number of bits of already encoded CUs of the ROI, I_r is the set of indexes of ROI CU that have not yet been coded, and $w_{\mathrm{CU}r}$ is the weight of the current CU of the ROI. Finally, the weight of each coding unit of index i is estimated by the MAD between the current unit p and the previous coded one p': $w_i = (\frac{1}{N} \sum_{j \in N} |p_j - p'_j|)^2$, where N is the number of pixels of the CU.

3.2. Region independent rate control models

Once the number of allocated bits for each CU is initialized, the QP is computed using the R-λ model. In our proposal the model of CUs from the ROI r is independent from the model of CUs of the non-ROI n. In fact, we have two models; in ROI, using the effective CU bits per pixel $R_{\mathrm{CU}r}$ of each unit,

$$\lambda_{\mathrm{CU}r} = \alpha_r\ R_{\mathrm{CU}r}^{\beta_r} \tag{9}$$

and for CUs from the non-ROI, using the effective CU bits per pixel $R_{\mathrm{CU}n}$,

$$\lambda_{\mathrm{CU}n} = \alpha_n\ R_{\mathrm{CU}n}^{\beta_n} \tag{10}$$

The model parameters are updated separately. For the ROI, the parameters α_r and β_r are updated referring to the original rate control algorithm [9], as follows:

$$\lambda_r = \alpha'_r\ R_r^{\beta'_r} \tag{11}$$

$$\alpha_r = \alpha'_r + 0.1\ (\ln \lambda'_r - \ln \lambda_r)\ \alpha'_r \tag{12}$$

$$\beta_r = \beta'_r + 0.05\ (\ln \lambda'_r - \ln \lambda_r)\ \ln R_r, \tag{13}$$

where α', β' and λ' are the old values of α, β and λ. In (11) and (13), R_r is the effective number of bits per pixel after encoding the unit. The same update process is used for the CUs of the non-ROI.

3.3. QP and λ variation

The last modification consists in considering new clipping ranges for λ and QP, at CU level. Thus, we can make different clipping for CUs of ROI and the other CUs. We allow a larger QP range than in the reference algorithm, to accommodate differences in quality between the ROI and the non-ROI. We define $\Delta\mathrm{QP}_{\mathrm{Pic}} > 2$ and $\Delta\mathrm{QP}_{\mathrm{CU}} > 1$ that guarantee

$$\mathrm{QP}_{\mathrm{Pic}} - \Delta\mathrm{QP}_{\mathrm{Pic}} \leq \mathrm{QP}_{\mathrm{CU}} \leq \mathrm{QP}_{\mathrm{Pic}} + \Delta\mathrm{QP}_{\mathrm{Pic}} \tag{14}$$

$$\mathrm{QP}_{\mathrm{CU}'} - \Delta\mathrm{QP}_{\mathrm{CU}} \leq \mathrm{QP}_{\mathrm{CU}} \leq \mathrm{QP}_{\mathrm{CU}'} + \Delta\mathrm{QP}_{\mathrm{CU}} \tag{15}$$

where $\mathrm{QP}_{\mathrm{CU}}$, $\mathrm{QP}_{\mathrm{Pic}}$ and $\mathrm{QP}_{\mathrm{CU}'}$ are respectively the QPs of the current CU, the current picture and the previously encoded CU.

4. EXPERIMENTAL RESULTS

We implemented the proposed rate control scheme in the HM.10 of HEVC [24]. To compute the ROI map, we used the Viola and Jones object detection method [11]. Then, a reference test is performed using the rate control algorithm described in [9]. This first test gives us the reference performance: the ratio K between ROI bit per pixel and non-ROI bit per pixel, the bit budget used for encoding each region and the PSNR of each region. Various sequences of different formats have been tested, with different bit partitioning, QP ranges and target bit rates. The examples reported here are "Johnny" and "Vidyo4" (1280x720 resolution with frame rate 60fps) [25] as the face detection algorithm gives the best face detection performances for these two sequences. The tested clipping ranges are $(\Delta \mathrm{QP}_{\mathrm{Pic}}, \mathrm{DeltaQP}_{\mathrm{CU}}) \in \{(2,1);(3,2)\}$ and the target bit rate evaluated is 128 kbps. Moreover, the tests are performed using a low delay B configuration with a GOP size equal to 4.

Johnny			Vidyo4		
K	Bit rate (kbps)	PSNR (dB)	K	Bit rate (kbps)	PSNR (dB)
5.49	127.96	37.15	3.67	128.02	34.07
7.69	127.93	36.85	5.01	127.86	33.95
8.50	127.98	36.68	5.68	127.77	33.83

Table 1. Global performance at 128kbps

Using hierarchical bit allocation at frame level and an LCU size equal to 64x64, the budget constraint is respected. Table 1 shows that our modifications to the rate control algorithm do not impair considerably the rate-distortion performance. At 128 kbps, we can increase the ratio compared to the reference $(K = 5.49)$ for "Johnny" and $(K = 3.67)$ for "Vidyo4" by keeping an output bit rate close to the assigned value. The overall PSNR decreasing slightly as our approach tends to alter the background quality which consists 87 % and 91 % of the frame, respectively, for "Johnny" and "Vidyo4".

Now we examine the quality of ROI and non-ROI for different bit rate ratios K. Table 2 shows the PSNR difference of each region between the proposed scheme and the original one. Overall, the bigger is K the better is the global quality of the ROI in the sequence and the lower is the PSNR of the non-ROI.

Johnny			Vidyo4		
K	ROI	non-ROI	K	ROI	non-ROI
7.69	0.56	-0.56	5.01	1.02	-0.25
8.50	0.57	-0.76	5.65	1.11	-0.39
9.00	0.66	-0.86	5.68	1.20	-0.41

Table 2. Δ PSNR (dB) for each region at 128kbps

We notice from Fig. 3 that the quality of the ROI is improved in all the GOPs while the quality of the non-ROI is decreased in all the GOPs. The performed tests show that the difference in quality of each region between the proposed scheme and the reference RC [9] is more important when K is bigger (as represented in Fig. 3). This means that our method allocates more bits to the ROI by improving its quality and respecting the overall bit rate constraint.

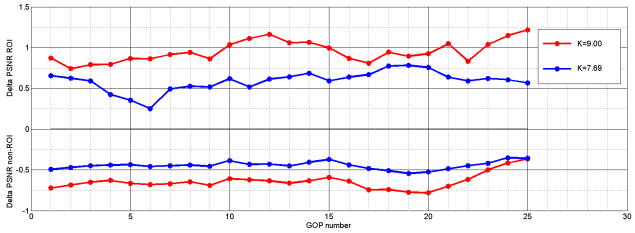
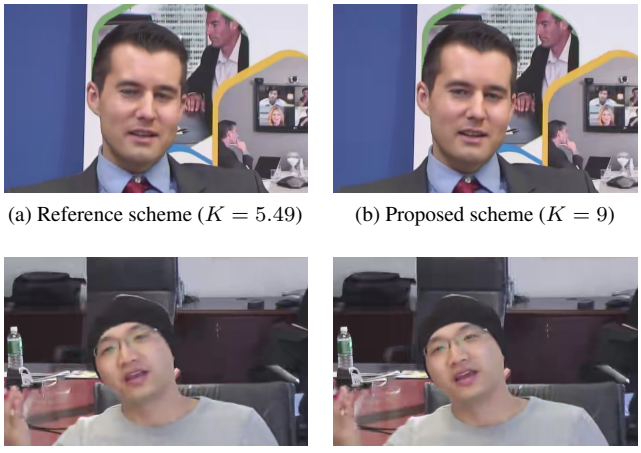


Fig. 3. ΔPSNR ROI and non-ROI for the last 25 GOPs of the sequence "Johnny"

Experimental results show both advantages in objective PSNR and subjective evaluation for ROI as represented in Fig. 4. We notice that, using our proposed scheme we can distinguish more details in the face and less artifacts, while the non-ROI does not present noticeable deterioration in visual quality.



(a) Reference scheme ($K = 5.49$) (b) Proposed scheme ($K = 9$)

(c) Reference scheme ($K = 3.67$) (d) Proposed scheme ($K = 5.65$)

Fig. 4. Subjective comparison of the face (ROI) for "Johnny" and "vidyo4" coded at 128kbps

5. CONCLUSION

In this paper, an ROI-based rate control for HEVC is proposed. The scheme achieves better visual quality in ROIs thanks to an independent processing at CU level of the two regions and a larger QP clipping range. The proposed algorithm shows better quality in ROI, while respecting the global bit rate constraint. This scheme is useful for videoconferencing systems by allowing a better representation of the facial expressions.

6. REFERENCES

[1] Z. Wu, S. Xie, K. Zhang, and R. Wu, "Rate Control in Video Coding," in *Recent Advances on Video Coding*, J. Del Ser Lorente, Ed. InTech, 2011, pp. 79–117.

[2] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Processing: Image Communication*, vol. 22, no. 1, pp. 19–38, Jan. 2007.

[3] L. Yang, L. Zhang, S. Ma, and D. Zhao, "A ROI quality adjustable rate control scheme for low bitrate video coding," in *Proceed. of Pict. Cod. Symp.* Ieee, May 2009, pp. 1–4.

[4] J. Chiang, C. Hsieh, G. Chang, F.-D. Jou, and W.-N. Lie, "Region-of-interest based rate control scheme with flexible quality on demand," in *Proceed. of IEEE Intern. Conf. on Multim. and Expo*, 2010, pp. 238–242.

[5] C.-Y. Wu and P.-C. Su, "A region-of-interest rate-control scheme for encoding traffic surveillance videos," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009, pp. 194–197.

[6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[7] H. Choi, J. Nam, J. Yoo, and D. Sim, "Rate control based on unified RQ model for HEVC," in *JCT-VC H0213, 8th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, San José, CA, USA, 2012.

[8] ——, "Improvement of the rate control based on pixel-based URQ model for HEVC," in *JCT-VC I0094, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2012.

[9] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," in *JCT-VC K0103, 11th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Shanghai, CN, 2012.

[10] B. Lee, M. Kim, and T. Nguyen, "A frame-level rate control scheme based on texture and non-texture rate models for High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–14, 2013.

[11] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR*, vol. 1, 2001, pp. I–511–I–518.

[12] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 74–90, Nov. 1998.

[13] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[14] S. Ma, J. Si, and S. Wang, "A study on the rate distortion modeling for High Efficiency Video Coding," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Sep. 2012, pp. 181–184.

[15] S. Yu and I. Ahmad, "A new rate control algorithm for MPEG-4 Video Coding," in *Proc. SPIE*, 2002.

[16] Z. Li, W. Gao, F. Pan, S. Ma, K. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. Lu, "Adaptive rate control for h.264," *Elsevier J. Vis. Comm. and Image Repres.*, vol. 17, no. 2, pp. 376–406, Apr. 2006.

[17] M. Naccari and F. Pereira, "Quadratic modeling rate control in the emerging HEVC standard," in *Proceed. of Pict. Cod. Symp.*, 2012, pp. 401–404.

[18] L. Sun, O. Au, and W. Dai, "An adaptive frame complexity based rate quantization model for intra-frame rate control of high efficiency video coding (hevc)," *APSIPA ASC*, pp. 1–6, 2012.

[19] J. Si, S. Ma, and W. Gao, "Adaptive rate control for HEVC," in *JCT-VC I0433, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2012.

[20] Y. Yoon, H. Kim, S.-h. Jung, and D. Jun, "A new rate control method for hierarchical video coding in HEVC," in *IEEE International symposium on Broadband Multimedia Systems and Broadcasting*, 2012.

[21] Z. He and S. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, 2002.

[22] Z. Ma, M. Xu, Y. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 5, pp. 671–682, 2012.

[23] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, 2005.

[24] *HEVC test model 10 (HM.10)*. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-10.1/

[25] F. Bossen, "Common test conditions and software reference configurations," in *JCT-VC L1100, 12th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013.