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 LTCI, UMR CNRS 5141

ABSTRACT

In this paper, we propose a new rate control scheme based on region of interest (ROI) for high efficiency video coding (HEVC). Our approach aims to allocate more bit resource for encoding the ROI while keeping the bit rate close to the assigned target value. This algorithm is developed for a videoconferencing system where the faces are automatically detected and each coding unit (CU) is classified in a ROI 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 rate and provides an improvement in ROI quality.

***Index Terms*—** HEVC, Rate control, ROI

1. INTRODUCTION

Rate control is an important tool that helps to deal with bit rate and compressed media quality fluctuations. Rate control methods have been widely studied and suitable schemes have been developed for specific applications [1]. It considers such challenging issues as resources availability, computational complexity, real-time [2].

One of the most interesting issues to focus on is the region of interest ROI. In video conferencing systems, surveillance and telemedicine the visual quality depends mainly on some important areas. 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 rate control 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 the input saliency map and the number of allocated bits of each region. For a video surveillance system, rate control 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 rate control model and are useful for H.264/AVC implementations [3] [4]. Meanwhile, the new HEVC standard has been finalized by ITU-T and ISO/IEC [6]. Thus, many have focused on rate control and developed new R-Q schemes for HEVC. 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 rate control 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 rate that processes separately the faces and the background. 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 a ROI map. The target bit rate is allocated for each region considering a fixed weight. Then, the QP of each CU is computed referring to the rate model of the corresponding region and the allocated bit budget. Finally, the proposed method deals with 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, 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, the conclusions are given in Section 5.

2. RELATED WORKS

The objective of rate control is to achieve a target bit rate as near as possible to a constant with a minimum quality distortion. Knowing that quantization consists on reducing the bit rate of the compressed video signal, the major role of rate control algorithms is to find for each sample the appropriate QP under the constraint $R_s(\mathrm{QP}) \leq \mathrm{R}_{\max}$. The fixed budget is $R_{\max}$ and $R_s(\mathrm{QP})$ is the number of bits of encoding the source sample s.

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

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

to achieve optimized mode decision of each CU. Using Lagrange multiplier λ_{MODE} in (1), the distortion $D(\mathrm{QP})$ is associated to the number of bits $R(\mathrm{QP})$ to evaluate the 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 are 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,

$$R(\mathrm{QP}) = \frac{C}{\mathrm{QP}} \tag{2}$$

where C is the model parameter. The quadratic model represented as

$$R(\mathrm{QP}) = \frac{C_1 \cdot \mathrm{MAD}}{\mathrm{QP}} + \frac{C_2 \cdot \mathrm{MAD}}{\mathrm{QP}^2} \tag{3}$$

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

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 difference (SAD) has been adopted in [19]. The MAD between original frame and reconstructed one has been introduced in [7] as represented in (3).

In a different way, the rate control has improved by considering representation in the ρ domain [20] like proposed in [21] and by taking into account other parameters such as the frame rate, e.g. the model proposed in [22].

The last rate distortion model in HEVC test software is the R-λ model represented as following

$$\lambda = \alpha \cdot \mathrm{bpp}^{\beta} \tag{4}$$

where α and β are the model parameters [9]. We note that this model defines a relationship between the rate in bit per pixel *bpp* and the Lagrange parameter λ which is used in RDO to decide the coding mode. Using 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 predict 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 in particular conditions. However, all rate control methods aim to allocate appropriate number of bits and determine the quantization parameter of each encoding unit. The complete R-λ rate control scheme in HEVC can be represented as follows,

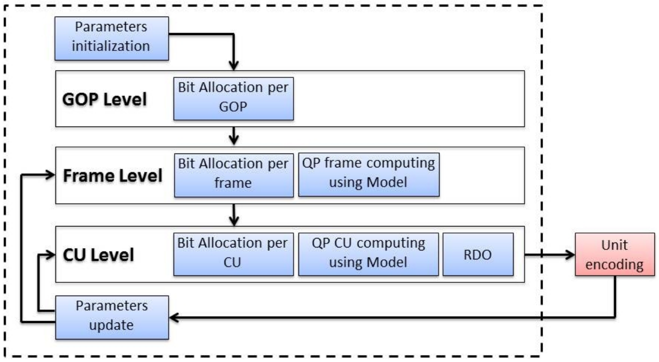


Fig. 1. Rate control scheme

As shown in figure 1, the controller operates in 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 picture.

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 frames hierarchical level. Then, the R-λ model is used to compute the frame QP.

iii. CU level: the process is divided into three parts. First, a 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 QP of the CU. λ and QP variation are clipped referring to a fixed range. Finally, the last step is the RDO 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 R-λ model for HEVC. The relationship between R and λ represented by (4) in section 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 in frame and CU levels and the computation of QP by the proposed model.

Figure 2 shows the proposed ROI-based rate control scheme. The first step consists in detecting the faces in the

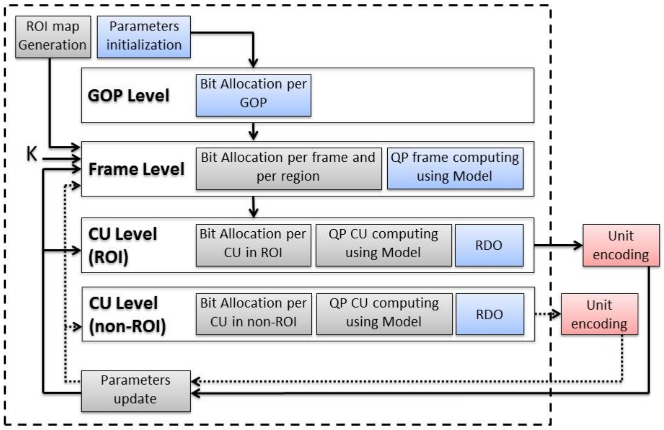


Fig. 2. ROI-based rate control scheme

scene and generating automatically a ROI map 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, frame budget is divided into two parts according to a fixed factor$K$ that is the ratio between the bit rate of the ROI and the bit rate of the rest of the frame (non-ROI). In 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 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, a positive constant K is selected. It represents the desired ratio between the ROI and non-ROI bit rates:

$$\mathrm{bpp}_{(\mathrm{r})} = K \cdot \mathrm{bpp}_{(\mathrm{n})} \tag{5}$$

Where, the index (r) denotes the ROI and (n) the non-ROI. We assume that the current allocated bit 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:

$$T_{\mathrm{Pic}} = T_{(r)} + T_{(n)} \tag{6}$$

$$T_{(n)} = \mathrm{bpp}_{(\mathrm{n})} \cdot M \cdot P_{(n)} \tag{7}$$

Where M is the total number of pixels of the frame and $P_{(n)}$ the percentage of non-ROI. From (5), (6) and (7), non-ROI bit rate $\mathrm{bpp}_{(\mathrm{n})}$ is computed as following:

$$\mathrm{bpp}_{(\mathrm{n})} = \frac{T_{\mathrm{Pic}}}{M \cdot (1 + P_{(r)} \cdot (K-1))} \tag{8}$$

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:

$$T_{\mathrm{CU(r)}} = \frac{T_{(r)} - T'_{(r)}}{\sum_{i \in I_{(r)}} w_{i(r)}} \cdot w_{\mathrm{CU(r)}} \tag{9}$$

Where $T'_{(r)}$ is the effective number of bit of already encoded CUs of the ROI. $I_{(r)}$ is the set of indexes of ROI CU that have not yet been coded. $w_{\mathrm{CU(r)}}$ 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 of the current unit p and the previous coded one p':

$$w_i = (\frac{1}{M} \sum_j |p_j - p'_j|)^2 \tag{10}$$

3.2. Region independent rate control models

Once the number of allocated bits for each CU is initialized, the QP is computed using R-λ model. In our proposal the model of CUs from the region of interest 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 $\mathrm{bpp}_{\mathrm{CU(r)}}$ of each unit,

$$\lambda_{\mathrm{CU(r)}} = \alpha_{(r)} \cdot \mathrm{bpp}_{\mathrm{CU(r)}}^{\beta_{(\mathrm{r})}} \tag{11}$$

and for CUs from NROI, using the effective CU bits per pixel $\mathrm{bpp}_{\mathrm{CU(n)}}$,

$$\lambda_{\mathrm{CU(n)}} = \alpha_{(n)} \cdot \mathrm{bpp}_{\mathrm{CU(n)}}^{\beta_{(\mathrm{n})}} \tag{12}$$

The models parameters are updated separately. For the ROI the parameters $\alpha_{(r)}$ and $\beta_{(r)}$ are updated referring to the original rate control algorithm [9], as following:

$$\lambda_{(r)} = \alpha'_{(r)} \cdot \mathrm{bpp}_{(\mathrm{r})}^{\beta'_{(\mathrm{r})}} \tag{13}$$

$$\alpha_{(r)} = \alpha'_{(r)} + 0.1 \cdot (\ln \lambda'_{(r)} - \ln \lambda_{(r)}) \cdot \alpha'_{(r)} \tag{14}$$

$$\beta_{(r)} = \beta'_{(r)} + 0.05 \cdot (\ln \lambda'_{(r)} - \ln \lambda_{(r)}) \cdot \ln \mathrm{bpp}_{(\mathrm{r})} \tag{15}$$

Where, α', β' and λ' are the old values of α, β and λ. In (13) and (15), $\mathrm{bpp}_{(\mathrm{r})}$ 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 on 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 larger QP range than the reference. We define $\Delta\mathrm{QP}_{\mathrm{Pic}} > 2$ and $\Delta\mathrm{QP}_{\mathrm{CU}} > 1$ that guarantees that

$$\mathrm{QP}_{\mathrm{Pic}} - \Delta\mathrm{QP}_{\mathrm{Pic}} \leq \mathrm{QP}_{\mathrm{CU}} \leq \mathrm{QP}_{\mathrm{Pic}} + \Delta\mathrm{QP}_{\mathrm{Pic}} \tag{16}$$

$$\mathrm{QP}_{\mathrm{CU'}} - \Delta\mathrm{QP}_{\mathrm{CU}} \leq \mathrm{QP}_{\mathrm{CU}} \leq \mathrm{QP}_{\mathrm{CU'}} + \Delta\mathrm{QP}_{\mathrm{CU}} \tag{17}$$

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 previous encoded CU.

4. EXPERIMENTAL RESULTS

We implemented the proposed rate control scheme on HM.10 of HEVC. To compute the ROI map, we used 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 reference performances: ratio between ROI bit rate and non-ROI bit rate K, bit budget used for encoding each region and PSNR of each region. Various sequences of different formats have been tested, with different bit partitioning and QP ranges. The example reported here is ”Johnny” (1280x720 pixels with frame rate 60fps). The tested clipping ranges are $\Delta\mathrm{QP}_{\mathrm{Pic}} = 3$ and $\Delta\mathrm{QP}_{\mathrm{CU}} = 2$. Moreover, the tests are performed using a low delay configuration [6] with a GOP size equal to 4.

K	Bit rate (kbps)	PSNR (dB)
3.77	101.13	32.64
4.05	101.24	32.61
5.27	101.91	32.41
5.89	101.72	32.24

Table 1. Global performances at 100kbps

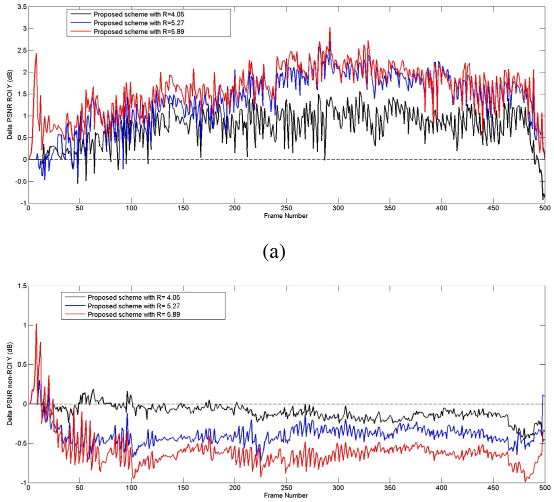
Table 1 shows that our modifications to the rate control algorithm do not impair the rate-distortion performances. At 100kbps, we can increase the ratio comparing to the reference $(K = 3.77)$ by keeping an output bit rate close to the assigned value. Moreover, the overall PSNR is practically the same as the reference encoder.

Now we examine the quality of ROI and non-ROI per different bit rate ratio 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.

K	Δ PSNR ROI (dB)	Δ PSNR non-ROI (dB)
3.88	0.55	-0.07
4.05	0.72	-0.11
5.27	1.35	-0.39
5.89	1.59	-0.60

Table 2. Global performances at 100kbps

In figure 3, we plot ΔPSNR of the ROI and ΔPSNR of the non-ROI per frame. We notice that the quality of the ROI is improving in all the frames while the quality of the non-ROI is decreasing in all the frames. The curves shows that for each region the difference in quality between the proposed scheme and the reference rate control [9] is more important when K is bigger. This means that our method leads to allocate more bit to the ROI by improving its quality and respecting the bit rate constraint.



(a)

(b)

Fig. 3. PSNR differences (ΔPSNR) between the proposed scheme and reference scheme for (a) ROI and (b) non-ROI

Experimental results show both advantages in objective PSNR and subjective evaluation for ROI as represented in figure 4. We notice that using our proposed scheme we can distinguish more details in the face and less artifacts. However, our ratio can not reach relatively big values. The non-ROI does not represent noticeable deterioration, which means that the background requires a minimum coding budget to keep the balance.

(a) (b)

Fig. 4. Subjective comparison for ”Johnny” coded at 100kbps (a)Reference scheme K=3.77 (b) Modified scheme K=5.89

5. CONCLUSION

In this paper, a ROI-based rate control for HEVC is proposed. The scheme achieves 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 bit rate constraint. This scheme is useful for videoconferencing systems to have a better encoding of the face 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.