Efficient Rate Control for Intra-Frame Coding in High Efficiency Video Coding

Feng Cen, Qianli Lu and Weisheng Xu
Department of Control Science and Engineering, Tongji University,
No. 1239 Siping Road, Shanghai, 200092, China

Keywords: HEVC, Intra-frame, Rate Control, Rate-quantization Model.

Abstract: In this paper, a coding tree unit (CTU)-level rate control (RC) scheme is proposed for intra-frame coding in the high efficiency video coding (HEVC). The CTU-level target bits are allocated based on the content complexity and the parameters of the Cauchy-based rate-quantization (R-Q) model of the current CTU is estimated according to the neighboring previously encoded CTUs. The proposed RC does not exploit any information of the adjacent frames such that it can inherently handle the initial frame and the scene change frames. The experimental results demonstrate the accurate rate estimation and stable video quality of the proposed RC scheme.

1 INTRODUCTION

Intra-frame is an important coding tool in the video coding. It can be not only used as the reference frame for the inter-frame coding in the video coding (Sullivan et al., 2012), but also independently used for the video coding and the still image coding in intra-only setup (Ku et al., 2006). Generally, the intra-predicted coding will generate much more bits than inter-predicted coding due to the lack of exploiting the temporal correlation between the frames. Hence, the accurate rate control (RC) of intra-frame coding is critical to consistently achieve good video quality for video coding under the limited bandwidth and buffer constraints.

Several RC schemes dedicated to intra-frame coding for H.264/AVC have been reported in the literature in recent years. Jing *et al.* (Jing et al., 2008) proposed a frame-level approach by utilizing a frame-level complexity measure based on the average gradient-per-pixel of the frame and the Cauchy-based rate estimation model. However, in their scheme, the estimation of the model parameters used for the current intra-frame depends on the preceding frames. Besides, how to accurately determine the initial quantization parameter (QP) for the sequence or a new scene is not addressed in their paper. Tsai *et al.* (Tsai and Chou, 2010) proposed a scene change aware RC approach for intra-only coding. They employed the Taylor-series approximation of Jing's model to circumvent the unreliable update of the non-stationary parameter, and proposed a rate estimation model for scene-transition frames. But, the performance of their RC algorithms highly relies on the precise of the scene change detection, which is still an open issue in video coding (Zeng et al., 2005), (Jing and Chau, 2006), (Yang et al., 2009). In addition, the parameters of their rate estimation models for the scene-transition frames are empirically obtained and may not be feasible for all practical scenarios.

As the latest video coding standard, the high efficiency video coding (HEVC) developed by the Joint Collaborative Team on Video Coding (JCT-VC) promises much higher compression efficiency than that possible with existing video coding standards (Ohm et al., 2012). Many new coding tools are developed for HEVC to improve the coding efficiency (Sullivan et al., 2012). Although many RC schemes have been proposed for HEVC (H. Choi and Sim, 2012) (Si et al., 2012) (Naccari and Pereira, 2012) (Li et al., 2012), but few of them are optimized for intra-frame coding. Recently, the R-λ model based scheme (Li et al., 2012) is adopted to the HEVC reference software HM14.0(Bossen et al.,), and on the basis of R-λ model, the Karczewicz et al. (Karczewicz and Wang, 2012) employed the sum of absolute transformed differences (SATD) to allocate the bit budget for intra-frame. However, the developed R-λ model is based on the conventional video quality measure PSNR. When the perceptual objective quality metrics, such as struc-

Cen F., Lu Q. and Xu W..
Efficient Rate Control for Intra-Frame Coding in High Efficiency Video Coding.
DOI: 10.5220/0005067100540059
In *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2014), pages 54-59
ISBN: 978-989-758-046-8

ture similarity index, are used to measure video quality, the merits of R-λ model are dubious. Furthermore, in (Karczewicz and Wang, 2012), the model parameters are updated according to previously encoded frames such that the scheme is not suited to handle the initial frame of a new scene.

In this paper, we present a novel coding tree unit (CTU)-level RC approach for intra-frame coding in HEVC. In the proposed scheme, the Cauchy-based R-Q model is applied to the CTU-level RC, the number of the target bits of each CTU is allocated according to the complexity measure of the CTU. To accurately initialize the model parameter, the first CTU of a frame is re-encoded on demand. The model parameters of the remaining CTUs in the frame are predicted by the model parameters of the preceding adjacent encoded CTUs of the current CTU. The advantages of the proposed scheme lie in that it can achieve accurate RC, inherently handle the scene transition frame regardless of the scene change detection and easily work together with the SSIM based rate-distortion optimization (RDO) algorithm (Cen et al., 2014) to achieve better perceptual video quality within the bandwidth constraint.

The rest of the paper is organized as follows. Section 2 introduces the proposed CTU-level RC approach. The experimental results are provided in section 3. Finally, section 4 concludes this paper.

2 PROPOSED CTU-LEVEL RC SCHEME

2.1 Complexity Measure

Instead of the macroblock, the core of the coding layer in HEVC is the CTU. To handle the variable block size in HEVC, the pixel-based complexity measure of the mth CTU in a frame is defined as the sum of the gradients of the pixels (SG) in the CTU, i.e.,

$$SG_m = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |I_{i,j} - I_{i+1,j}| + |I_{i,j} - I_{i,j+1}|, \qquad (1)$$

where $I_{i,j}$ denotes the luminance of the pixel at the location of (i, j) , and M and N are the horizontal and vertical dimensions of the luma samples of the CTU, respectively. It is well known that the number of generated bits r and SG have a linear relationship at the frame level (Jing et al., 2008). After analyzing plenty of frames from various sequences, we observe that such a linear relationship is still approximately held even at the CTU level, as shown in Fig. 1. Here, r_m denotes the number of the generated bits of the mth CTU.

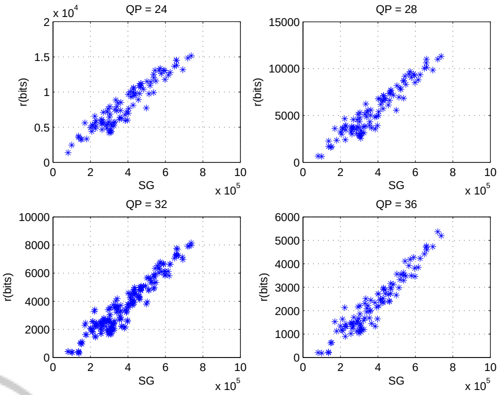


Figure 1: Scatter plots of r_m versus SG_m for all the CTUs in the fifth frame of *PartyScene* (*WVGA*) sequence under $QP = 24, 28, 32, 36$. Apparently, there exists a linear relationship between r_m and SG_m for coding with fixed QP.

2.2 CTU-level Bit Allocation

Owing to the fact that r_m is approximately proportional to SG_m, we allocate the number of the target bits of the mth CTU according to its complexity measure. Let T_F be the number of the target bits of the current frame. Then, the number of the target bits of the mth CTU, t_m, is given by

$$t_m = (T_F - \Sigma_{i=1}^{m-1} r_i) \frac{SG_m}{SG_F - \Sigma_{i=1}^{m-1} SG_i}, \qquad (2)$$

where SG_F denotes the frame-level SG, which is the sum of all the SG_ms in the frame.

2.3 R-Q Model

Analogous to the frame-level Cauchy-based rate estimation model in (Jing et al., 2008), we formulate r_m with

$$r_m = SG_m a_m Q_m^{b_m}, \qquad (3)$$

where a_m and b_m are model parameters for the mth CTU, and Q_m is the quantization step size for the mth CTU. For convenience, we define the complexity-normalized r_m as

$$\Upsilon_m = \frac{r_m}{SG_m}. \qquad (4)$$

The examples of $\Upsilon_m - Q$ are shown in Fig. 2.

Theoretically, a_m and b_m can be determined according to the Cauchy probability density function of the transform coefficients (Altunbasak and Kamaci, 2004). However, since only after encoding can the actual distribution of the transform coefficients in a CTU be obtained, these two model parameters should be predicted. We have analyzed a large number of

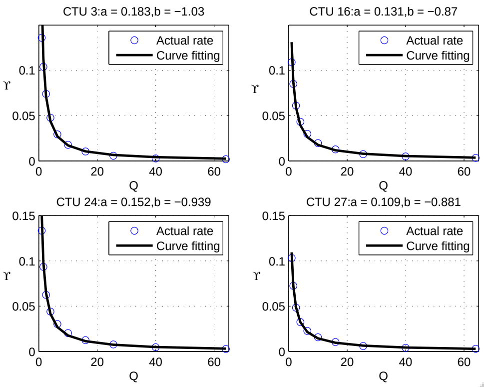


Figure 2: Curve fitting results between Υ_m and Q for four randomly selected CTUs with the size of 64×64 in the fifth frame of *BasketballPass* sequence. The QP values range from 4 to 40, at an increment of 4.

frames from sequences with various texture characteristics, and found that the optimal b_m is fallen in the range of -1.05 to -0.85 and the variation of b_m has little impact on the accuracy of the rate estimation. Hence, for simplicity, we fix b_m at a moderate value -0.9 for the model used in our experiments and omit the subscript m hereinafter.

Now, only a_m is left to be determined. The parameter a_m is estimated as follows. For the first CTU in a intra-frame, a_1 is firstly initialized with a constant. In our experiments, we set a_1 to 0.142. If the relative error between target and generated bits, which is defined as

$$\Delta r_1 = \frac{|t_1 - r_1|}{t_1}, \tag{5}$$

is greater than a threshold τ (empirically, we set $\tau = 0.3$), then a_1 is updated with the real Υ_1 as follows,

$$a_1 = \frac{\Upsilon_1}{Q_1^b}, \tag{6}$$

and the first CTU is re-encoded with the new a_1.

For the following CTUs in the frame, a_m is predicted by using the actual complexity-normalized generated bits, actual model parameter a_m and quantization step sizes of the encoded CTUs on the left, top and top-left positions of the current CTU, i.e.,

$$a_m = \begin{cases} \beta a_{m-1} + (1-\beta)\frac{\Upsilon_{m-1}}{Q_{m-1}^b}, & m < W+1 \\ \beta a_{m-1} + (1-\beta)\frac{\Upsilon_{m-W}}{Q_{m-W}^b}, & (m\%W) == 0 \\ \beta a_{m-1} + (1-\beta)(\frac{\Upsilon_{m-1}}{Q_{m-1}^b}\frac{\Upsilon_{m-W}}{Q_{m-W}^b}\frac{\Upsilon_{m-W-1}}{Q_{m-W-1}^b})^{\frac{1}{3}}, & \text{otherwise} \end{cases}, \tag{7}$$

where W is the number of CTUs in the horizontal dimension of the current frame, and β is a weight factor. In our experiments, we set $\beta = 0.2$. After estimating a_m, we can determine the QP of the current CTU by (3) according to the target bits.

2.4 RC Scheme

In summary, we illustrate the proposed CTU-level RC scheme in Fig. 3. Note that in order to keep the smoothness of the visual quality, we restrict the maximum QP change between consecutive CTUs in the frame to 2.

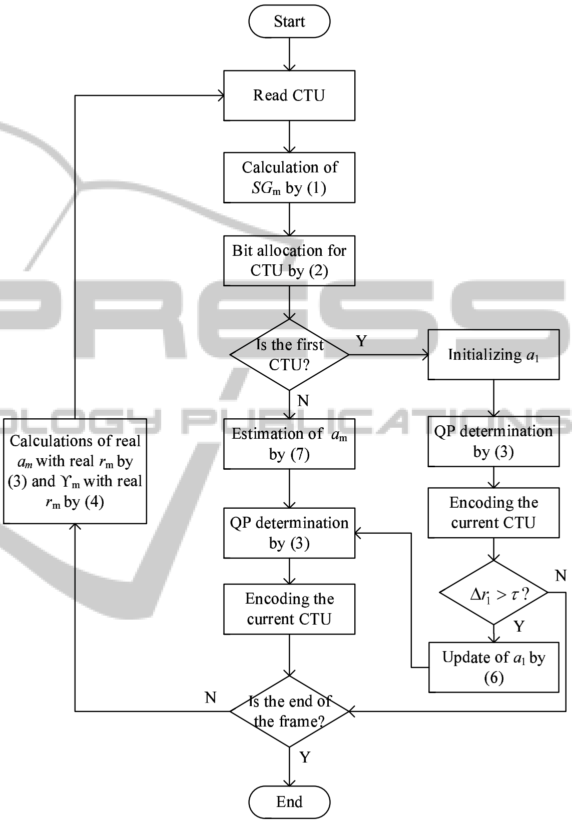


Figure 3: Flowchart of the proposed RC scheme for the coding of an intra-frame.

3 EXPERIMENTAL RESULTS

3.1 Rate Estimation Accuracy

In HEVC, the CTU has the variable size. Fig. 4 compares the actual and estimated Υ_ms for the first two rows of the CTUs in the seventh frame of *BlowingBubbles* sequences. It can be observed that the estimated Υ_m is very close to the actual Υ_m and can track the change of the actual Υ_m. Moreover, we can observe that the estimation accuracy degrades for the small size of the CTU. Although the proposed RC scheme can be extended to coding unit (CU)-level, we do not recommend to apply it to the luma block size smaller than 16×16.

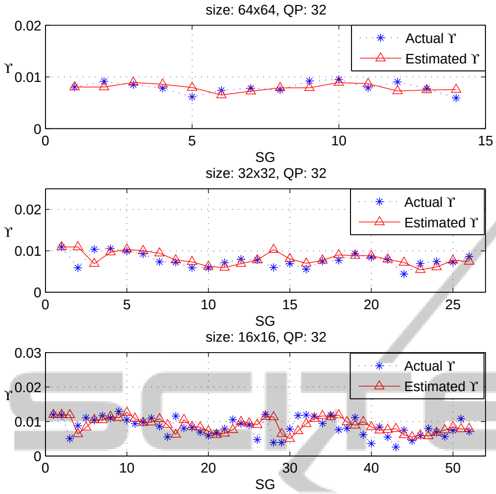


Figure 4: Comparison of the actual and estimated ϒ's for the first two rows of the CTUs in the seventh frame of *BlowingBubbles* sequences. The QP is fixed to 32. The CTU sizes from top to bottom are 64x64, 32x32 and 16x16, respectively.

3.2 RC Performance

To evaluate the performance, we integrated our proposed RC approach into the HM 8.0 (Bossen et al.,). The CTU size is set to 64×64, the maximum allowed size in HEVC. In our experiments, a commonly used simple and efficient bit allocation strategy (Tsai and Chou, 2010) that assigns the same bits for the remaining frames is adopted for the frame-level target bit budget $T_{F,l}$, i.e.,

$$T_{F,l} = \frac{R_r}{K_r}, \tag{8}$$

where R_r and K_r are the available bits for the remaining frames and the number of the remaining frames, respectively.

The first group of experiments is conducted by using the first 150 frames of eight test sequences selected from Class A∼D specified in the common test conditions (Bossen, 2011). All test sequences are encoded in constant bit-rate (CBR) setting with the intra-only high efficient configuration, and the target bit-rate for each test sequence is set to the average bit-rate generated by HM8.0 with a fixed QP. Table 1 tabulates the average frame-level bit mismatch ratio $M\%$, peak mismatch ratio $PM\%$ and $\Delta PSNR$ for all the test sequences. $M\%$, $PM\%$ and $\Delta PSNR$ are calculated as follows,

$$M\% = \frac{1}{K}\sum_{l=1}^{K}\frac{|T_{F,l}-R_{F,l}|}{T_{F,l}} \times 100\%, \tag{9}$$

$$PM\% = \max_{l=1,\ldots,K}\left(\frac{|T_{F,l}-R_{F,l}|}{T_{F,l}} \times 100\%\right), \tag{10}$$

$$\Delta PSNR = \overline{PSNR} - \overline{PSNR_o}, \tag{11}$$

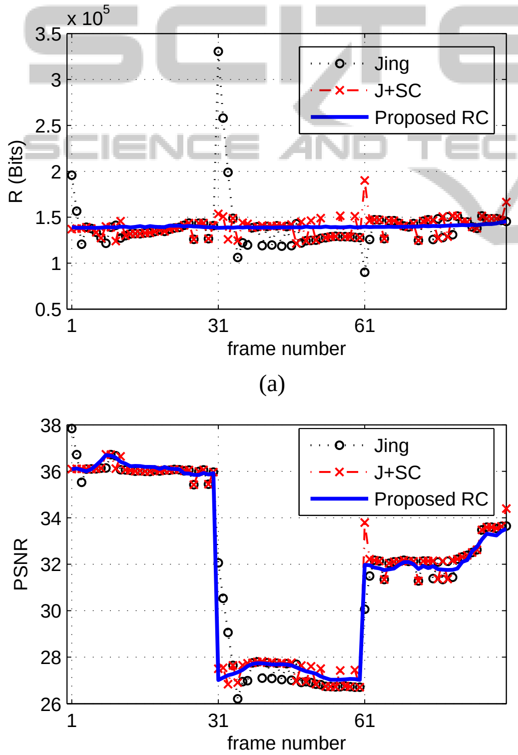
respectively, where K is the number of frames in the sequence, and $R_{F,l}$ is the generated bits for the lth frame. $\overline{PSNR}$ and $\overline{PSNR_o}$ are the average PSNRs obtained by using the proposed RC approach and the fixed QP approach in HM 8.0, respectively. In Table 1, the QP indicates that the target bit-rate is obtained by using HM8.0 with the fixed QP equal to the corresponding value. From Table 1, we can observe that $M\%$ and $PM\%$ are quite small. Meanwhile, the average $PSNR$ for the proposed RC approach is also quite close to that for the the HM8.0 encoder with the fixed QP. These results demonstrate that the proposed RC approach can accurately control the bit-rate without any sacrifice in video quality.

Table 1: Performance of the proposed RC scheme.

Sequence	QP	$M\%$	$PM\%$	$\Delta PSNR(dB)$
PeopleOnStreet	24	1.51	1.73	0.11
(1600P)	32	1.46	1.60	0.06
Traffic	24	1.27	1.43	0.04
(1600P)	32	0.97	1.14	0.04
ParkScene	24	1.07	1.61	-0.09
(1080P)	32	0.43	0.98	-0.08
BasketballDrive	24	0.60	1.47	-0.20
(1080P)	32	0.83	2.40	-0.25
BasketballDrill	24	1.01	1.28	0.12
(WVGA)	32	0.52	1.98	0.03
PartyScene	24	1.05	1.85	0.13
(WVGA)	32	1.16	1.56	0.06
BlowingBubbles	24	0.90	1.54	0.04
(QWVGA)	32	0.82	1.11	0.00
BQSquare	24	1.03	1.39	0.09
(QWVGA)	32	0.62	1.32	0.04
Avg.		0.95	1.52	0.0261

To demonstrate the merits of the proposed RC scheme to handle scene changes, we use a combination sequence *Combo* (*BasketballDrill-PartyScene-BQMall*, WVGA), which is generated by cascading the first 30 frames from the three test sequences, for the experiments. The prominent H.264/AVC intra-frame RC approaches, Jing's RC approach (Jing et al., 2008) (denoted by "Jing's") and Jing's RC approach with two-pass encoding for each scene-transition frame (denoted by "J+SC"), are implemented to HM8.0 for comparison. In "J+SC", the two-pass encoding technique can improve the RC performance at the scene-transition frames. This is because although an inaccurate initial QP may be used in the first-pass encoding, the outcome of the first-pass encoding is used to improve the accuracy of the frame-level rate estimation model, and thus, a better QP can be determined with the improved rate estimation model in the seconde-pass encoding. Furthermore, in the experiments with ”Jing’s” and ”J+SC” schemes, we assume that there is a perfect scene change detection, although it is not true for the stat of the art automatical scene change detection algorithm. So, instead of using an automatical scene change detection, we manually point out the scene transition frame for the experiments with ”Jing’s” and ”J+SC” schemes.

Fig. 5(a) and Fig. 5(b) show the comparisons of the generated bits and PSNRs frame by frame, respectively. As it can be seen from Fig. 5, the bit-rate of the proposed RC scheme is more stable than those of the other two approaches and the PSNR of the proposed RC scheme is smoother between frames in the same scene than those of the other two approaches. Particularly, even without any indication about the scene transition frames, at the scene transition frame the proposed RC scheme works better than the other two approaches. This is because the proposed RC scheme does not rely on the scene change detection. Furthermore, from table 2, we can observe that the $M\%$ is reduced by 95.4% and 84.3% and the PSNR fluctuation (variance) in the scene for different scenes reduced by 10.2% – 76.0%, respectively. Here, The PSNR standard deviation, Dev, is defined as



(a)

(b)

Figure 5: Comparisons of (a) generated bits and (b) PSNR for the *Combo* sequence. The scene changes occur at the 31st and 61st frames The frame rate is 30 f/s, and the target bandwidth is 4096 kbit/s.

$$Dev = \sqrt{\frac{1}{K}\sum_{l=1}^{K}(PSNR_l - \overline{PSNR})^2}, \qquad (12)$$

where $PSNR_l$ denotes the PSNR of the lth frame in the scene. The improvement of the proposed RC scheme is attributed to the CTU-level rate adjustment and the accurate estimation of the model parameter and the generated bits. Since there is at most one CTU that is needed to be re-encoded, the proposed RC scheme has low delay and computational complexity and is particular suitable for the live video coding.

Table 2: Performance comparison of the Jing’s, J+SC and the proposed RC schemes.

	Jing’s	J+SC	Proposed RC
$M\%$	9.7%	6.01%	0.94%
PSNR	31.84	31.75	31.88
Dev: 1st scene	0.445	0.263	0.226
Dev: 2nd scene	1.165	0.443	0.279
Dev: 3rd scene	0.659	0.593	0.532

4 CONCLUSION

In this paper, a simple and efficient CTU-level RC approach for intra-frame coding is presented. The proposed RC approach has the advantage in maintaining the stability of the bit-rate and PSNR, especially for the sequence containing multiple scene changes. Moreover, due to the pixel-based complexity measure, the proposed RC approach can be easily extended to slice-level or CU-level.

ACKNOWLEDGEMENTS

The authors would like to thank the support of the National Natural Science Foundation of China No. 60972035.

REFERENCES

Altunbasak, Y. and Kamaci, N. (2004). An analysis of the DCT coefficient distribution with the H.264 video coder. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 177–180.

Bossen, F. (2011). Common test conditions and software reference configurations. In *ITU-T SG16 Contribution, JCTVC-E700*, Daegu.

Bossen, F., Flynn, D., and Suehring, K. HM reference software. [Online]. *https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/*.

Cen, F., Lu, Q., and Xu, W. (2014). Ssim based rate-distortion optimization for intra-only coding in hevc. In *Consumer Electronics (ICCE), 2014 IEEE International Conference on*, pages 17–18. IEEE.

H. Choi, J. Nam, J. Y. and Sim, D. (2012). Rate control based on unified RQ model for HEVC. In *ITU-T SG16 Contribution, JCTVC-H0213*, San José.

Jing, X., Chau, L., and Siu, W. (2008). Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control. *IEEE Signal Processing Letters*, 15:373–376.

Jing, X. and Chau, L.-P. (2006). A novel intra-rate estimation method for H.264 rate control. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pages 4–pp. IEEE.

Karczewicz, M. and Wang, X. (2012). Intra frame rate control based on SATD. *ITU-T SG16 Contribution, JCTVC-M0257, Incheon*, pages 1–5.

Ku, C.-W., Cheng, C.-C., Yu, G.-S., Tsai, M.-C., and Chang, T.-S. (2006). A high-definition H.264/AVC intra-frame codec IP for digital video and still camera applications. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(8):917–928.

Li, B., Li, H., Li, L., and Zhang, J. (2012). Rate control by R-lambda model for HEVC. *ITU-T SG16 Contribution, JCTVC-K0103, Shanghai*, pages 1–11.

Naccari, M. and Pereira, F. (2012). Quadratic modeling rate control in the emerging HEVC standard. In *Picture Coding Symposium (PCS), 2012*, pages 401–404. IEEE.

Ohm, J., Sullivan, G. J., Schwarz, H., Tan, T. K., and Wiegand, T. (2012). Comparison of the coding efficiency of video coding standardsłincluding high efficiency video coding (HEVC). *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1669–1684.

Si, J., Ma, S., Zhang, X., and Gao, W. (2012). Adaptive rate control for high efficiency video coding. In *Visual Communications and Image Processing (VCIP), 2012 IEEE*, pages 1–6. IEEE.

Sullivan, G. J., Ohm, J., Han, W.-J., and Wiegand, T. (2012). Overview of the high efficiency video coding (HEVC) standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 22(12):1649–1668.

Tsai, W. and Chou, T. (2010). Scene change aware intra-frame rate control for H.264/AVC. *Circuits and Systems for Video Technology, IEEE Transactions on*, (99):1–1.

Yang, M., Serrano, J. C., and Grecos, C. (2009). MPEG-7 descriptors based shot detection and adaptive initial quantization parameter estimation for the H.264/AVC. *IEEE Transactions on Broadcasting*, 55(2):165–177.

Zeng, W., Du, J., Gao, W., and Huang, Q. (2005). Robust moving object segmentation on H.264/AVC compressed video using the block-based MRF model. *Real-Time Imaging*, 11(4):290–299.