Background Error Propagation Model Based RDO for Coding Surveillance and Conference Videos

Jian Xiong[1], Xianzhong Long[1], Ran Shi[2], Miaohui Wang[3], Jie Yang[1], Yunyi Li[1], and Guan Gui[1(✉)]

[1] Nanjing University of Posts and Telecommunications, Nanjing, China
{jxiong,guiguan}@njupt.edu.cn
[2] Nanjing University of Science and Technology, Nanjing, China
[3] Shenzhen University, Shenzhen, China

Abstract. Surveillance and conference videos have become increasingly important in our daily life, which brings a huge amount of video data. Existing coding standards were originally designed for generic video contents. The backgrounds are generally static in the surveillance and conference videos. The background coding errors will propagate to the subsequent frames in coding the videos. In this paper, a background error propagation (BEP) model based Rate Distortion Optimization (RDO) scheme in HEVC is proposed for the surveillance and conference videos. Firstly, the global RDO scheme is proposed to efficiently exploit the background error propagation. Secondly, a BEP model is studied to express the linear relationship between the distortion of the first frame and that of its subsequent frames. Based on the BEP model, enhanced frames are proposed to be coded with a small quantization parameter (QP) offset so as to improve the global performance. Thirdly, a bi-exponential decay model is proposed to investigate the variation of the error propagation ratio as the frame order increased. Based on the decay model, a periodical optimization scheme is presented by deploying the enhanced frames periodically. Experimental results show that the proposed algorithm achieves 11.15% bit-rate reductions on average under the low delay condition.

Keywords: HEVC · Video coding · Error propagation · Surveillance · Background modeling

This workRecently, video was supported in part by the National Natural Science Foundation of China Grant (No. 61701258, No. 61701310), Natural Science Foundation of Jiangsu Province Grant (No. BK20170906, No. BK20150856), Natural Science Foundation of Jiangsu Higher Education Institutions Grant (No. 17KJB510044), Jiangsu Specially Appointed Professor Grant (RK002STP16001), Innovation and Entrepreneurship of Jiangsu High-level Talent Grant (CZ0010617002), and "1311 Talent Plan" of Nanjing University of Posts and Telecommunications, High-Level Talent Startup Grant of Nanjing University of Posts and Telecommunications (XK0010915026).

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 504–513, 2019.
https://doi.org/10.1007/978-3-030-19086-6_56

1 Introduction

Recently, video surveillance and conference systems are becoming more and more prevalent in our daily life. As it was reported by IDC [2], surveillance videos will grow to 5,800 exabytes by 2020. In the face of the explosive growth of surveillance videos, how to effectively compress the videos has become a significant big challenge.

The state-of-the-art video coding standards, such as H.264/AVC [5] and High Efficiency Video Coding (HEVC) [3,4] are widely used to compress the surveillance and conference videos. In these methods, coding tools including intra prediction, motion estimation (ME), transformation, and quantization are employed to remove the redundancy. Rate-distortion optimization (RDO) technology is adopted to select the optimal coding modes and parameters. However, these methods were originally designed for generic video contents. Different from the generic videos, the surveillance videos generally acquired with static cameras. In these videos, the backgrounds are static and the motion patterns are generally simple. The coding errors in the background regions may propagate to the subsequent frames. This characteristic was not fully studied in the traditional coding methods.

Many efforts have been made to investigate more efficient methods for coding the surveillance and conference videos. by modeling background frames [6–9]. In [8], the HEVC hierarchical prediction is optimized with background modeling for surveillance and conference video coding. The background picture is generated and encoded as the long-term reference frame. In [7], a background modeling based adaptive prediction is proposed for surveillance video coding. The long-term redundancy is reduced by predicting on generated background frames. Adaptive prediction methods are employed for different coding blocks. The background generation is performed on basis of the frames, and the generated background is updated for each group of pictures (GOP). In [9] a selective background difference coding method is proposed on basis of macro-block (MB) level. Two ways are selected to code the macro-blocks. One is coding the original MB, and the other is directly coding the difference between the MB and the corresponding background. A block-based background modeling method is proposed for surveillance video coding [6]. In this scheme, background generation and updating is conducted based on coding units (CUs) but not frames and is performed for every frame but not a whole GOP. However, in these methods, only one generated picture cannot model the periodical backgrounds efficiently. The generated background may get worse as the frame distance increases. Furthermore, the block-based background modeling methods may aggravate the block artifacts between the foreground regions and background regions.

In the recent works, background modeling based schemes are proposed to exploit the frame dependency. However, the background error propagation characteristic is not fully studied. In this paper, a background error propagation (BEP) model based global RDO scheme is proposed for surveillance and conference video coding. The BEP model is presented to describe the linear relationship between the distortion of the frames. In this model, a concept of propagation ratio is proposed to describe how is the distortion of one frame influenced by its previous frames. Based on the BEP model, enhanced frames are presented to be coded with a small quantization parameter (QP) offset. Furthermore, a bi-exponential decay model is proposed to express the variation of the propagation ratio as the frame order increased. Based on the decay model, the periodical optimization scheme is presented by periodically coding enhanced frames. Experiments are tested on surveillance and conference videos. Experimental results show the efficiency of the proposed method.

The rest of the paper is organized as follows. An overview of HEVC RDO technology is presented in Sect. 2. The proposed BEP model based global RDO method is given in Sect. 3. Experiments are provided in Sect. 4 to validate the efficiency of the proposed method. Finally, we draw some concluding remarks in Sect. 5.

2 Overview of Rate Distortion Optimization

RDO technology is widely used in the block-based hybrid coding standards, such as H.264/AVC and HEVC. In these standards, there are various coding modes and parameters which can be employed to code the blocks. RDO is employed to select the optimal coding modes and parameters. The fundamental problem of RDO is to minimize the coding distortion with a bit consumption constraint. The constraint problem can be converted into an unconstrained problem by introducing a Lagrangian multiplier. It can be expressed by

$$\min J = D + \lambda \cdot R, \tag{1}$$

where the symbols R and D denote the coding bits and the corresponding coding distortion. The parameter λ denotes the Lagrangian multiplier. There is a trade-off between the distortion and coding bits. A proper Lagrangian multiplier will lead to an optimal balance. The default λ is obtained from the input QP value, which is expressed by,

$$\lambda = fac \cdot \frac{(qp - 12)}{3}, \tag{2}$$

where fac is the QP factor, qp is the input QP value.

3 Proposed Method

3.1 Global Rate Distortion Optimization

In the traditional coding scheme, RDO technology is independently employed to code each CU. However, in practical applications, when we try to code a video sequence, the main goal is to code all the frames with the optimal rate-distortion balance. There is a strong frame dependency in the consecutive frames, especially for the surveillance and conference videos. Thus, a global optimization scheme is more applicable for coding all the consecutive frames than the independent scheme. The global RDO scheme is given by,

$$\min J = \sum_{f=1}^{k} D_f + \lambda \cdot \sum_{f=1}^{k} R_f. \tag{3}$$

where k is the coded frame number. The symbols D_f and R_f denote the distortion and the corresponding coding bits of the fth $(f = 1, 2, \ldots, k)$ frame, respectively. In contrast with the traditional RDO technology, the global optimization scheme considers all the consecutive frames but not only one CU.

3.2 Background Error Propagation Model

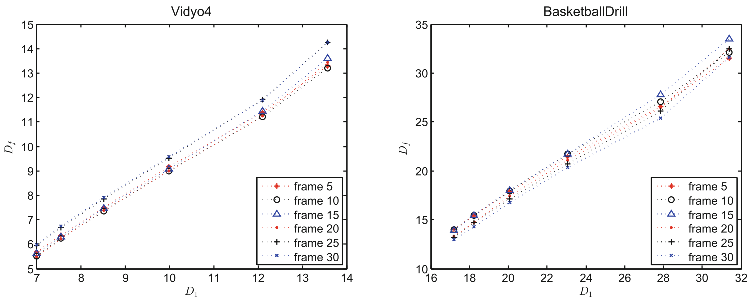
Generally, because of the prediction coding scheme in existing coding standards, coding errors may propagate from the previous frame to the subsequent frames. The frame dependency is not being well used in the existing optimization.

In surveillance videos, the backgrounds are static and the motion patterns are generally simple. Let's consider k co-located CUs in the temporal consecutive frames. On one hand, the original co-located background pixels in temporal consecutive CUs are reasonable to be considered as the same. This is expressed by, $P_{1,j} = P_{2,j} = P_{3,j} = \ldots = P_{k,j}$, where $j = 1, 2, \ldots, N^2$ denote the pixel locations in CUs with size $N \times N$. On the other hand, since CUs in background regions are generally encoded with the skip mode, the reconstructed pixels are considered to be approximately equal with each other, denoted as $P^d_{1,j} \approx P^d_{2,j} \approx P^d_{3,j} \approx \ldots \approx P^d_{k,j}$. Therefore, for $i = 1, 2, \ldots, k$, the relationship between the CU distortion can be written as

$$\begin{aligned} SSD_i &= \sum_{j=1}^{N^2} (P_{i,j} - P^d_{i,j})^2 \\ &\approx \sum_{j=1}^{N^2} (P_{1,j} - P^d_{1,j})^2 \\ &\approx SSD_1, \end{aligned} \tag{4}$$

where SSD_i denotes the sum of squared differences for CU i. That is, in the background regions, the distortion of consecutive CUs is approximately equal with that of the first frame.

Based on the background error propagation characteristic as in (4), it can be concluded that the distortion of subsequent frames is significantly influenced by that of the first frame. Experiments are conducted to study the relationship between the distortion of the first frame and that of the subsequent frames. In the experiments, the coding structure is IPPP. Except for the first inter frame, all the inter frames are coded with a fixed QP value set as 35. The first inter frame is coded with QP value varies from 23 to 33 with an interval 2. As shown in Fig. 1, the X-axis represents the distortion of first inter frame, and the y-axis represents



(a) (b)

Fig. 1. The relationship between the distortion of the first frame and that of its subsequent frames. With the sequences: (a) Vidyo4, (b) BasketballDrill.

the distortion of the subsequent frame (such as frame 5, 10, 15, 20, 25, and 30). It can be observed that the distortion of the subsequent frames is highly influenced by that of the first frame, and there is a strong linear relationship between the distortion. It is reasonable to assume a linear model as

$$D_f = r_f \cdot D_1 + b_f, \tag{5}$$

where b_f is a bias term, r_f is a parameter which represents the error propagation ratio. The linear model is named as background error propagation model. The error propagation ratio in the model describes how is the distortion of one frame influenced by its previous frames.

3.3 BEP-Based Rate Distortion Optimization

Based on the background error propagation model, the global RDO shown in (3) can be rewritten as

$$\min(D_1 \cdot \sum_{f=1}^{k} r_f + \sum_{f=1}^{k} b_f + \lambda R_1 + \lambda \cdot \sum_{i=2}^{k} R_i). \tag{6}$$

It can be observed that, since the subsequent frames are significantly influenced by the first frame, improving the coding performance of the first frame will make the overall coding performance to be enhanced. Thus, we try to improve the coding performance of the first frame. On the other hand, the bit-rate of each frame is nearly independent. That is to say, the optimal coding performance of the total k frames can be obtained by setting the following derivative to 0. It is expressed by,

$$\frac{\partial(D_1 \cdot \sum_{f=1}^{k} r_f + \lambda R_1)}{\partial R_1} = 0. \tag{7}$$

Thus, the lambda multiplier can be solved as,

$$\lambda_1 = -\frac{\partial D_1}{\partial R_1} = \frac{\lambda}{\sum_{f=1}^{k} r_f}, \tag{8}$$

where λ_1 denotes the lambda multiplier of the first frame.

From (2), we can have

$$qp = 3log_2(\frac{\lambda}{fac}) + 12. \tag{9}$$

Combine with (8), the adjusted QP value of the first frame can be calculated by λ_1 as

$$qp' = qp - 3log_2(\sum_{f=1}^{k} r_f). \tag{10}$$

For convenience, we use s to denote the summation of error ratios, i.e., $s = \sum_{f=1}^{k} r_f$. The coding performance of the first frame can be improved by coding it with a small QP offset, which is given by

$$\Delta Q = round(-3log_2(s)), \tag{11}$$

where ΔQ denotes the QP offset. The frame which is coded with the small QP offset is named as the *enhanced frame*.

3.4 Bi-exponential Decay Model

Since the QP offset depends on the summation of error ratios s, experiments are conducted to investigate the propagation ratio of the BEP model. In the experiments, two sets of tests are performed on the first 60 frames. The first set is named as the *anchor set*, in which the coding structure is the Low-Delay P setting, and the quantization parameter (QP) is set to 32. In order to investigate the error propagation characteristic, another set of tests is performed by setting QP value as 1 for encoding the first inter frame (approximately lossless coding). It should be noticed that the QP values for encoding the other frames are not changed. This set of tests is named as the *improved set*.

The distortion is measured in terms of mean square errors (MSE). For the anchor set, the distortion of frame f is denoted as D_f. For the improved set, the distortion of frame f is denoted as $\widetilde{D_f}$. By comparing the anchor with the improved set, the error increment of each frame f can be calculated as $\Delta D_f = D_f - \widetilde{D_f}$. The error propagation ratio between frame f and the first inter frame can be measured as

$$r_f = \Delta D_f / \Delta D_1. \tag{12}$$

Figure 2 shows the error propagation ratio of the P-frames. The x-axis represents the frame order number. The y-axis represents the error propagation ratio of each frame. It indicates that there is a strong biexponential decay relationship between the error increment and the frame order number, which can be expressed by

$$r_f = \eta_1 \cdot {\eta_2}^f + \eta_3 \cdot {\eta_4}^f. \tag{13}$$

The symbols η_1, η_2, η_3, and η_4 are the model parameters. The decay model shows that the error propagation ratio decreases as the frame order number increases. Equation (4) shows that the background pixels have a strong error propagation characteristic. However, even in surveillance videos, not all the pixels are in background regions. Foreground regions with motion objects are common in the videos. Thus, the decay model is reasonable because as the frame order number increases, fewer pixels have the error propagation property.

In addition, we evaluate the fitting goodness of biexponential decay model. As shown in Fig. 2, the average R-square value (denoted as R^2) is 0.976. That is, the biexponential decay model has high accuracy in modeling the downtrend of the error increments.

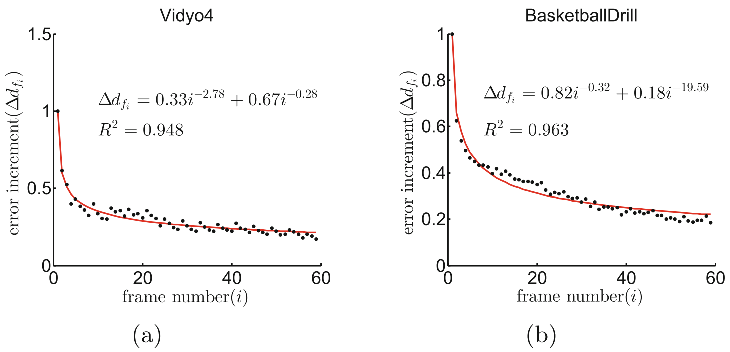


Fig. 2. The bi-exponential decay model. With the sequences: (a) Vidyo4, (b) BasketballDrill.

3.5 Implementation

As it is indicated in the decay model, the propagation ratio decreases as the frame order number increases. That is to say, the influence of the first frame on far-distance frames is small. It is necessary to set a new enhanced frame for the far-distance frames. Therefore, the enhanced frames are necessary to be deployed periodically. The interval between two enhanced frames is defined as an optimization period.

Figure 3 shows the proposed periodical RDO scheme. In this figure, the yellow bar denotes an I frame, and the other bars are P frames. The numbers in the gray box are the QP offsets. The red bars denote the enhanced frames coded with a QP offset as ΔQ. There is an optimization between two enhanced frames.

As shown in Fig. 2, all the propagation ratios become small and converge at frame 60. Thus, every 60 frames is coded as an optimization period, i.e., $k = 60$.

Table 1. Summation of the error propagation ratios when the optimization period is set to 60.

Sequences	s
Vidyo4	17.45
Vidyo3	16.01
Traffic	17.71
BasketballDrill	18.63
Average	17.45

That is, the first frame of each optimization period is coding the QP offset ΔQ. Table 1 shows the sum of error propagation ratios when the optimization period is set to 60. It indicates that the average value of s is 17.45, and most of the values are close to the average. By employing the average s in (11), we obtain the QP offset as -12, i.e., $\Delta Q = -12$.

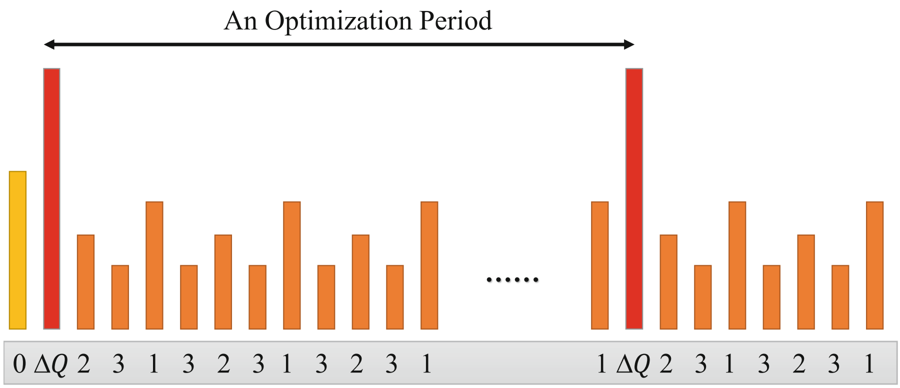


Fig. 3. The background error propagation model based global RDO scheme. The yellow bar denotes the I frame, and the other bars are P frames. The numbers in the gray box are the QP offsets. The red bars denote the enhanced frames coding with a QP offset as ΔQ. (Color figure online)

4 Experimental Results

The experiments were performed on a PC with an Intel (R) 3.60 GHz processor, 16 Gb RAM. The performance of the proposed method is evaluated in terms of the change of the Bjontegaard Delta bit-rate (BD-BR) and Bjontegaard Delta Peak Signal to Noise Ratio (BD-PSNR) [1]. The proposed method is integrated on the HEVC reference software, HM16.0[1]. The performance gain is obtained by comparing the proposed method with the reference software.

[1] https://hevc.hhi.fraunhofer.de/svn/svnHEVCSoftware/tags/HM-16.0/

Table 2. R-D performance improvements of the proposed method compared with the hevc default scheme.

Sequences	BD-BR (%)				BD-PNSR (dB)			
	Y	U	V	YUV	Y	U	V	YUV
Vidyo1	−8.90	−31.21	−32.55	−9.74	0.25	0.51	0.69	0.26
Vidyo3	−7.53	−38.51	−47.54	−9.88	0.19	0.67	1.33	0.25
Vidyo4	−12.01	−34.22	−32.74	−12.89	0.30	0.73	0.76	0.32
Traffic	−0.48	−15.22	−23.53	−1.34	0.01	0.27	0.45	0.03
BasketballDrill	−10.33	−9.74	−9.19	−15.33	0.43	0.41	0.50	0.59
Johnny	−9.07	−34.83	−31.69	−10.70	0.16	0.69	0.64	0.19
KristenAndSara	−11.28	−32.69	−32.83	−13.05	0.31	0.79	0.77	0.35
FourPeople	−13.70	−32.97	−31.71	−14.87	0.44	0.90	0.88	0.47
Average	−9.20	−28.31	−29.89	−11.15	0.26	0.64	0.76	0.31

In the experiment, 8 sequences captured with static cameras were tested since the proposed method is aiming at static background videos. The experiment setting is the low delay ("encoder_lowdelay_P_main"). In order to cover different ranges of qualities and bit-rates, the proposed method is tested with a groups of QP values including 22, 27, 32, and 37.

Table 2 shows the R-D performance of the proposed method tested on the common setting (on the first group). It can be observed that the proposed can significantly improve the coding performance. The average BD-BR reductions over the anchor are 9.20%, 28.31%, and 29.89% on Y, U, and V components, respectively. The corresponding BD-PSNR increments are 0.26 dB, 0.64 dB, and 0.76 dB on Y, U and V components, respectively. The weighted BD-BR reduction (denoted as YUV BD-BR) and BD-PSNR increment (denoted as YUV BD-PSNR) of all components are 11.15% and 0.31 dB, respectively. It indicates that the proposed algorithm significantly outperforms the default HEVC method. Furthermore, especially for the sequences with a large proportion of static background (such as FourPeople, KristenAndSara, and Johnny), the performance gain is larger than that of the sequence with a small proportion of static background (such as Traffic). It indicates that the proposed method is better suited to the static background.

5 Conclusion

In this paper, a BEP model based global RDO method in HEVC is proposed for surveillance and conference videos. The proposed method is different from the default RDO scheme, in which each CU is optimized independently. Since the backgrounds are generally static in the surveillance and conference videos, the R-D performance of the long-term frames is optimized globally in the proposed method, in which the background error propagation can be efficiently exploited. Two models are presented to study the characteristics of background error propagation. The first one is the linear BEP model, which describes the linear relationship between the distortion of the first frame and that of subsequent frames. Based on the BEP model, enhanced frames coded with a small QP offset is deployed to improve the global performance. The second one is the bi-exponential decay model, which expresses the variation of the error propagation ratio as the frame order increased. Based on the decay model, a periodical optimization scheme is presented, i.e., the enhanced frames are periodically deployed. Experimental results show that the proposed algorithm achieves an average 11.15% bit-rate reduction on YUV components (YUV BD-BR) for the low delay setting.

References

1. Bjontegaard, G.: Calculation of average PSNR differences between RD curves. No. ITU-T SC16/Q6, VCEG-M33, Austin, USA, April 2001
2. Gantz, J., Reinsel, D.: The IDC digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east (2012). https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf
3. Han, W.J., et al.: Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools. IEEE Trans. Circuits Syst. Video Technol. **20**(12), 1709–1720 (2010). https://doi.org/10.1109/TCSVT.2010.2092612
4. Sullivan, G., Ohm, J., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol. **22**(12), 1649–1668 (2012). https://doi.org/10.1109/TCSVT.2012.2221191
5. Wiegand, T., Sullivan, G., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. **13**(7), 560–576 (2003). https://doi.org/10.1109/TCSVT.2003.815165
6. Yin, L., Hu, R., Chen, S., Xiao, J., Hu, J.: A block-based background model for surveillance video coding. In: 2015 Data Compression Conference, pp. 476–476, April 2015. https://doi.org/10.1109/DCC.2015.49
7. Zhang, X., Huang, T., Tian, Y., Gao, W.: Background-modeling-based adaptive prediction for surveillance video coding. IEEE Trans. Image Process. **23**(2), 769–784 (2014). https://doi.org/10.1109/TIP.2013.2294549
8. Zhang, X., Tian, Y., Huang, T., Dong, S., Gao, W.: Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling. IEEE Trans. Image Processing **23**(10), 4511–4526 (2014). https://doi.org/10.1109/TIP.2014.2352036
9. Zhang, X., Tian, Y., Liang, L., Huang, T., Gao, W.: Macro-block-level selective background difference coding for surveillance video. In: 2012 IEEE International Conference on Multimedia and Expo, pp. 1067–1072, July 2012. https://doi.org/10.1109/ICME.2012.136