IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Background Error Propagation Model based RDO in HEVC for Surveillance and Conference Video Coding

JIAN XIONG[1], (Member, IEEE), XIANZHONG LONG[2], RAN SHI[3], MIAOHUI WANG[4], JIE YANG[1], AND GUAN GUI.[1], (Senior Member, IEEE)

[1]College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China, 210003 (e-mail: jxiong@njupt.edu.cn,jyang@njupt.edu.cn and guiguan@njupt.edu.cn)
[2]School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, China, 210023, (e-mail: lxz@njupt.edu.cn)
[3]School of Computer Science, Nanjing University of Science and Technology, Nanjing, China, 210094, (e-mail: rshi@njust.edu.cn)
[4]School of Information Engineering, Shenzhen University, Shenzhen, China, 518060, (e-mail: mhwang@szu.edu.cn)

Corresponding authors: Jian Xiong (e-mail: jxiong@njupt.edu.cn) and Guan Gui (e-mail: guiguan@njupt.edu.cn).

This work was supported in part by the National Natural Science Foundation of China Grant (No. 61701258, No. 61701310, No. 61801219), Natural Science Foundation of Jiangsu Province Grant (No. BK20170906, No. BK20150856), Natural Science Foundation of Jiangsu Higher Education Institutions Grant (No. 17KJB510044), NUPTSF Grant (NY215030), 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 (NUPT), High-Level Talent Startup Grant of NUPT (XK0010915026).

ABSTRACT The emerging High Efficiency Video Coding (HEVC) Standard has significantly improved the compression performance in comparison with its predecessor H.264/AVC. However, it was 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 R-D performance of the long-term frames is optimized globally. The global RDO scheme can 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) so as to improve the global performance. Thirdly, a 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. Experiments are tested on the surveillance and conference videos. The results show that the proposed algorithm achieves significant performance improvement.

INDEX TERMS HEVC, video coding, error propagation, surveillance, background modeling

I. INTRODUCTION

RECENTLY video surveillance and conference systems are becoming increasingly more common in our daily life. As it was reported by IDC [1], surveillance videos will grow to 5,800 exabytes by 2020. In the face of the explosive growth of surveillance video, how to effectively compress it has become a significantly big challenge.

The state-of-the-art video coding standards, such as H.264/AVC [2] and High Efficiency Video Coding (HEVC) [3]–[5] 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 [6]–[8]. 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. The first class is background modeling based coding methods [9]–[12], which generate background frames for reducing the long-term redundancy. In [9], 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 [10], 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 [11]. 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 [12]. 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, 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.

The second class employs background and foreground information for efficient coding [13]–[17]. In [13], a background proportion based adaptive Lagrangian multiplier selection method is proposed for surveillance video coding. The Lagrangian multiplier in RDO is changed by a lambda factor parameter, which is trained based on a feature of background proportion. In [14], a bit allocation algorithm is proposed based on the background and foreground information for surveillance video coding. The Largest Coding Units (LCUs) are classified into foreground LCUs and background LCUs, and then different bit allocation schemes are employed for the two classes respectively. In [15], a foreground-based scheme is proposed for low bitrate surveillance video coding. The foreground frame is obtained by segmentation. Foreground frame based ME is more accurate for foreground prediction. In [17], A background-foreground division based ME method is proposed for surveillance video coding. The prediction units (PUs) are classified into foreground PUs and background PUs. Two different Motion estimation strategies are adopted for the two classes respectively.

In addition to the above approaches, there are many other methods studied on surveillance video coding from various aspects [18]–[22]. In [18], a knowledge-based coding scheme is proposed for encoding multisource surveillance video data. It tries to reduce the global redundancy of foreground objects across multiple cameras. In [19], a vehicle library based surveillance video coding method is presented to remove the global redundancy. In [20], the dynamic textural synthesis method is proposed to compress surveillance videos, in which a histogram of the motion direction based method is proposed to detect dynamic textural. The low-rank and sparse decomposition are applied to compress surveillance videos by reducing the strong temporal redundancy [16], [21]. Different from the common schemes which target on objective/subjective quality [23]–[27], a novel framework of surveillance video coding is proposed for improving the coding efficiency towards intelligent analysis performance [22].

In the recent works, background modeling and background information 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). Furthermore, a 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 Section II. The proposed BEP model based global RDO method is given in Section III. Experiments are provided in Section IV to validate the efficiency of the proposed method. Finally, we draw some concluding remarks in Section V.

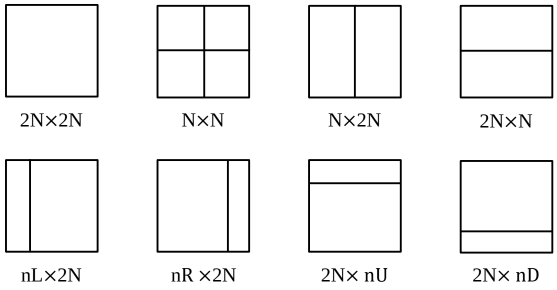


FIGURE 1. The partition modes of Prediction Unit (PU).

II. 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. Take HEVC as an example, a video frame is divided into a set of Coding Tree Units (CTUs). CTU is a quad-tree structure, in which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2879329, IEEE Access

the nodes are called as Coding Units (CUs). The sizes of CU include 8×8, 16×16, 32×32 and 64×64 [28]. Furthermore, Prediction Unit (PU) and Transform Unit (TU) are also introduced in HEVC. Each CU can have multiple PUs and TUs. As shown in Fig 1, there are up to 8 PU partition modes. TUs are also allowed to have various sizes from 4×4 to 32×32. The various partitioning of CU, PU, and TU improve the compression efficiency of HEVC. 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. It can be expressed by,

$$\min D \ \text{ s.t. } \ R < R_c, \tag{1}$$

where R_c is the bit number constraint. The symbols R and D denote the coding bits and the corresponding coding distortion.

The constraint problem can be converted into an unconstrained problem by introducing a Lagrangian multiplier. Eq. (1) can be rewritten as

$$\min J = D + \lambda \cdot R, \tag{2}$$

where 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,

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

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

III. PROPOSED METHOD

A. 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 to coding all the consecutive frames than the independent scheme. The global RDO scheme is given by,

$$\min J = \sum_{f=1}^{k} D_f + \lambda \cdot \sum_{f=1}^{k} R_f. \tag{4}$$

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 ($f = 1, 2, ..., k$) frame, respectively. In contrast with the traditional RDO technology, the global optimization scheme considers all the consecutive frames but not only one CU.

B. 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, $P_{1,j} = P_{2,j} = P_{3,j} = ... = P_{k,j}$, where $j = 1, 2, ..., N^2$ 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 $P^d_{1,j} \approx P^d_{2,j} \approx P^d_{3,j} \approx ... \approx P^d_{k,j}$. Therefore, for $i = 1, 2, ..., k$, the relationship between the CU distortion can be written as

$$\begin{aligned} SSD_i &= \sum_{j=1}^{N^2} (P_{i,j} - P^d_{i,j})^2 \\ &\approx \sum_{j=1}^{N^2} (P_{1,j} - P^d_{1,j})^2 \\ &\approx SSD_1, \end{aligned} \tag{5}$$

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 (5), 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. 2, the x-axis represents the distortion of the first inter frame, and the y-axis represents 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

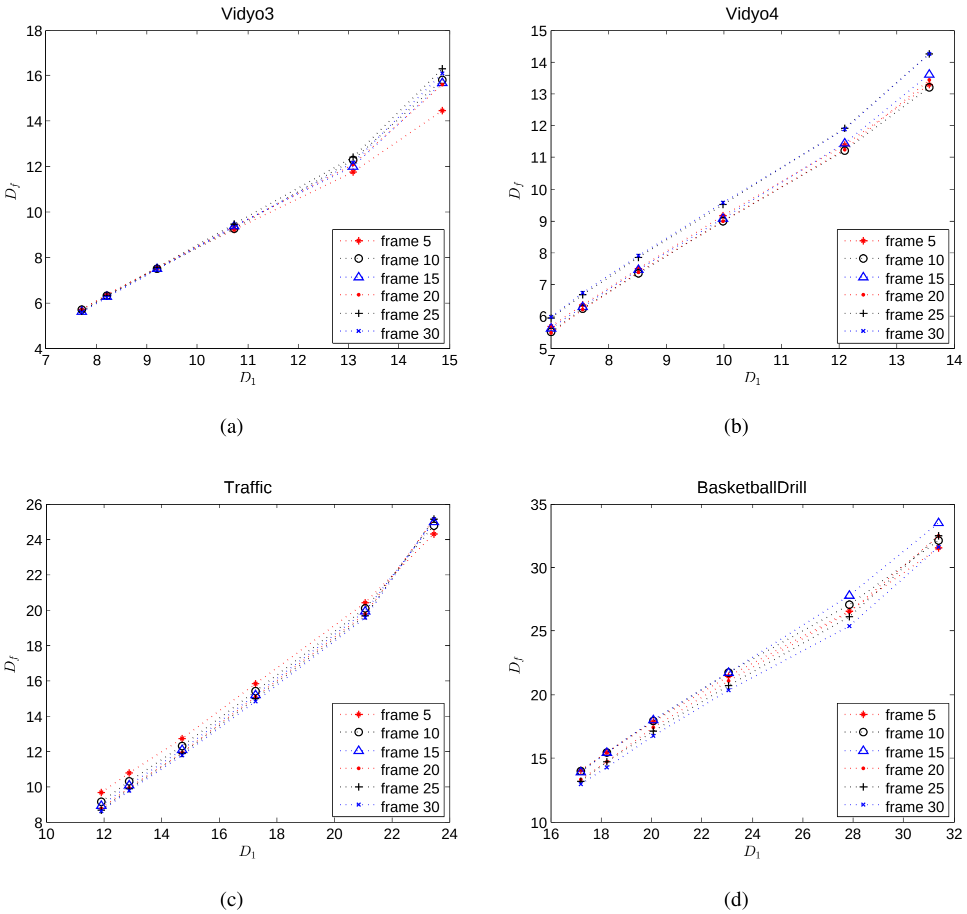
$$D_f = r_f \cdot D_1 + b_f, \tag{6}$$

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.

C. BEP-BASED RATE DISTORTION OPTIMIZATION

Based on the background error propagation model, the global RDO shown in (4) can be rewritten as

$$\min(D_1 \cdot \sum_{f=1}^{k} r_f + \sum_{f=1}^{k} b_f + \lambda R_1 + \lambda \cdot \sum_{i=2}^{k} R_i). \tag{7}$$



(a)

(b)

(c)

(d)

FIGURE 2. The relationship between the distortion of the first frame and that of its subsequent frames. With the sequences: (a) Vidyo3, (b) Vidyo4, (c) Traffic, and (d) BasketballDrill.

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,

$$\frac{\partial(D_1 \cdot \sum_{f=1}^{k} r_f + \lambda R_1)}{\partial R_1} = 0. \tag{8}$$

Thus, the lambda multiplier can be solved as,

$$\lambda_1 = -\frac{\partial D_1}{\partial R_1} = \frac{\lambda}{\sum_{f=1}^{k} r_f}, \tag{9}$$

where λ_1 denotes the lambda multiplier of the first frame.

From (3), we can have

$$qp = 3log_2(\frac{\lambda}{fac}) + 12. \tag{10}$$

Combine with (9), the adjusted QP value of the first frame can be calculated by λ_1 as

$$qp' = qp - 3log_2(\sum_{f=1}^{k} r_f). \tag{11}$$

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, which is given by

$$\Delta Q = round(-3log_2(s)), \tag{12}$$

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

D. ERROR PROPAGATION RATIO 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 of four sequences (Traffic, Vidyo3, Vidyo4, and BasketballDrill). The first set is named as the anchor set, in which the coding structure is the Low-Delay P setting, and the quantization parameter

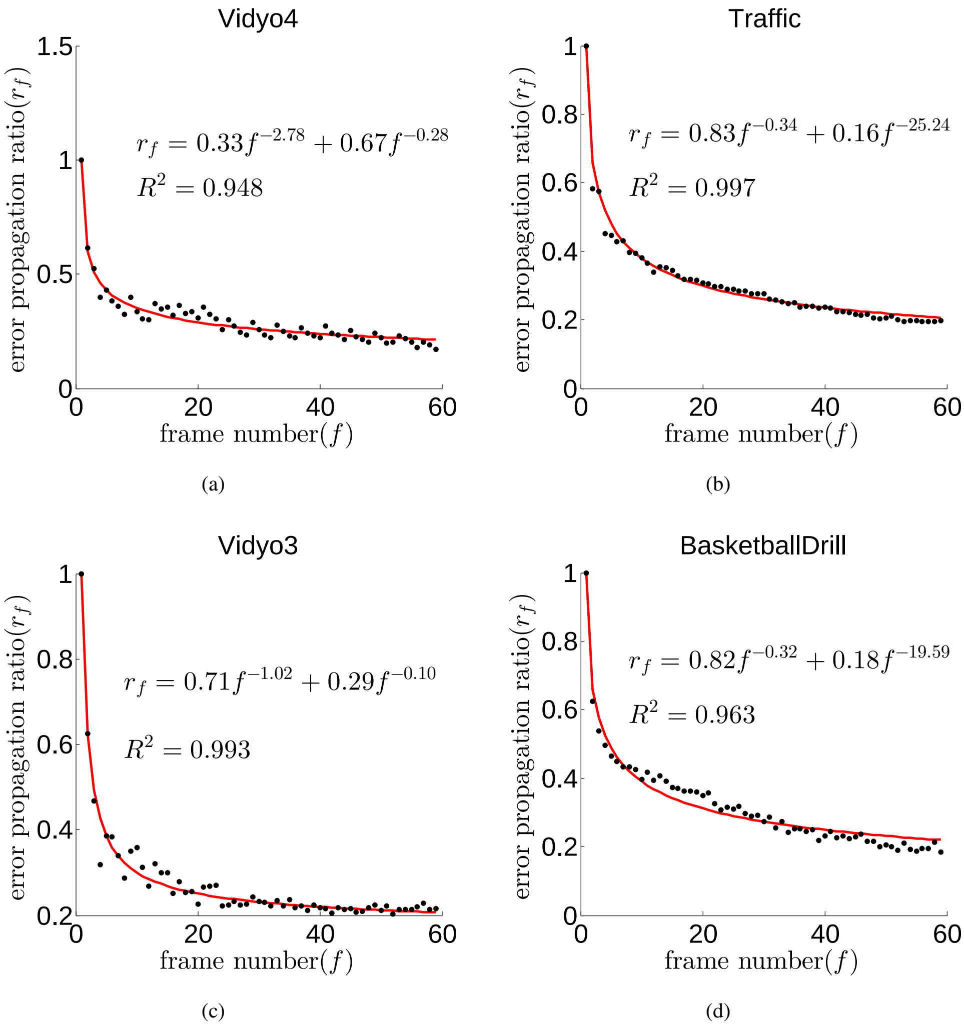


FIGURE 3. The error propagation ratio decay model. With the sequences: (a) Vidyo4, (b) Traffic, (c) Vidyo3, and (d) BasketballDrill.

(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

$$r_f = \Delta D_f / \Delta D_1. \tag{13}$$

Fig. 3 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 decay relationship between the error increment and the frame order number, which can be expressed by

$$r_f = \eta_1 \cdot f^{\eta_2} + \eta_3 \cdot f^{\eta_4}. \tag{14}$$

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. Eq. (5) 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 the decay model. As shown in Fig. 3, the average R-square value (denoted as R^2) is 0.976. That is, the decay model has high accuracy in modeling the downtrend of the error increments.

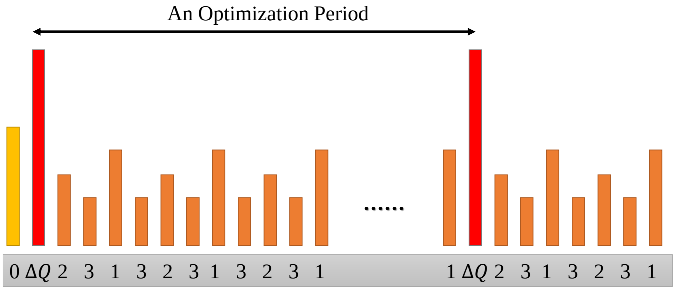


FIGURE 4. The background error propagation model based global RDO scheme. The yellow bar denotes an I frame, and the other bars are P frames. The numbers in the gray box are the QP offsets. The red bars denote the enhanced frames coding with a QP offset as ΔQ.

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

E. 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.

Fig. 4 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. 3, all the propagation ratios become small and converge at frame 60. Thus, every 60 frames are coded as an optimization period, i.e., $k = 60$. 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 (12), we obtain the QP offset as -12, i.e., $\Delta Q = -12$.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

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) [29]. The performance gain is obtained by comparing the proposed method with the reference software. The proposed method is integrated on the reference software, HM16.0[1].

Since the proposed method focuses on videos with static backgrounds, 12 video sequences including conference videos and surveillance videos were tested in the experiments. Six conference video sequences, including Vidyo1, Vidyo3, Vidyo4, Johnny, FourPeople and Kristen&Sara, were originally used to evaluate the performance of HEVC. Six surveillance video sequences, including Bank, Crossroad, Office, Overbridge, Intersection and Mainroad are obtained from the PKU-SVD-A dataset [9], [10], [30]. The resolutions of the sequences include standard definition (SD) and high definitions (HD).

The low delay setting ("encoder_lowdelay_P_main") is adopted in the experiments. In order to cover different ranges of qualities and bit-rates, the proposed method is tested with two groups of QP values. The QP values in the first group are the common test QPs including 22, 27, 32, and 37. The QP values in the second group are set to 29, 32, 35, and 38, which are employed to validate the proposed method on low bit-rate applications.

B. PERFORMANCE IMPROVEMENT ON HEVC DEFAULT SCHEME

Table 2 shows the R-D performance of the proposed method tested on the first group of QPs. It can be observed that the proposed method can significantly improve the coding performance. When tested on the conference videos, the average BD-BR reductions over the anchor are 10.72%, 34.64%, and 35.30% on Y, U, and V components, respectively. The corresponding BD-PSNR increments are 0.28dB, 0.76dB, and 0.87dB 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 12.28% and 0.31dB, respectively. When tested on the surveillance videos, the average BD-BR reductions are 14.39%, 49.80%, and 47.30%, on Y, U, V components, respectively. The corresponding BD-PSNR increments are 0.33dB, 1.14dB, and 1.11dB on Y, U, and V components, respectively. The YUV BD-BR reduction and BD-PSNR increment are 15.42% and 0.35dB, respectively. It indicates that the proposed algorithm significantly outperforms the default HEVC method for coding both the conference videos and the surveillance videos.

The results also show that, the average BD-BR reductions tested on the surveillance videos are larger than that of the conference videos. The reason is, most of the surveillance videos have larger proportions of static backgrounds than that of the conference videos. Furthermore, especially for the sequences with a large proportion of static background (such as Bank and Overbridge), the performance gain is larger than that of the sequence with a small proportion of static background (such as Crossroad and Office). It indicates that the proposed method is better suited to the static background.

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

TABLE 2. R-D PERFORMANCE IMPROVEMENTS OF THE PROPOSED METHOD COMPARED WITH THE HEVC DEFAULT SCHEME (ON COMMON SETTING).

Sequences		BD-BR (%)				BD-PNSR (dB)			
		Y	U	V	YUV	Y	U	V	YUV
Conference Videos	Vidyo1-720p	-8.90	-31.21	-32.55	-9.74	0.25	0.51	0.69	0.26
	Vidyo3-720p	-7.53	-38.51	-47.54	-9.88	0.19	0.67	1.33	0.25
	Vidyo4-720p	-12.01	-34.22	-32.74	-12.89	0.30	0.73	0.76	0.32
	Johnny-720p	-9.07	-34.83	-31.69	-10.70	0.16	0.69	0.64	0.19
	KristenAndSara-720p	-11.28	-32.69	-32.83	-13.05	0.31	0.79	0.77	0.35
	FourPeople-720p	-13.70	-32.97	-31.71	-14.87	0.44	0.90	0.88	0.47
	Average	**-10.72**	**-34.64**	**-35.30**	**-12.28**	**0.28**	**0.76**	**0.87**	**0.31**
Surveillance Videos	Bank-sd	-20.90	-55.91	-54.34	-27.21	0.25	1.27	1.56	0.35
	Crossroad-sd	-3.71	-51.51	-49.08	-3.75	0.17	1.24	1.00	0.16
	Office-sd	-7.04	-59.86	-58.36	-10.55	0.16	1.40	1.05	0.19
	Overbridge-sd	-27.37	-56.79	-55.26	-31.54	0.63	1.96	2.08	0.74
	Intersection-hd	-13.49	-24.22	-22.55	-12.26	0.40	0.50	0.48	0.35
	Mainroad-hd	-13.84	-50.52	-44.23	-10.38	0.38	0.47	0.46	0.31
	Average	**-14.39**	**-49.80**	**-47.30**	**-15.42**	**0.33**	**1.14**	**1.11**	**0.35**

TABLE 3. R-D PERFORMANCE IMPROVEMENTS OF THE PROPOSED METHOD COMPARED WITH THE HEVC DEFAULT SCHEME (ON LOW BIT-RATE SETTING).

Sequences		BD-BR (%)				BD-PNSR (dB)			
		Y	U	V	YUV	Y	U	V	YUV
Conference Videos	Vidyo1-720p	-9.62	-26.06	-27.02	-9.90	0.38	0.60	0.73	0.37
	Vidyo3-720p	-9.51	-31.47	-33.47	-10.34	0.37	0.68	1.25	0.39
	Vidyo4-720p	-12.43	-24.74	-21.70	-12.42	0.45	0.67	0.65	0.44
	Johnny-720p	-12.02	-19.04	-23.75	-12.26	0.40	0.62	0.69	0.40
	KristenAndSara-720p	-12.82	-25.06	-23.53	-13.38	0.49	0.78	0.75	0.50
	FourPeople-720p	-16.45	-25.54	-22.51	-16.49	0.73	0.84	0.79	0.71
	Average	**-12.14**	**-25.32**	**-25.33**	**-12.46**	**0.47**	**0.70**	**0.81**	**0.47**
Surveillance Videos	Bank-sd	-33.86	-34.48	-34.52	-34.04	1.31	1.09	1.29	1.28
	Crossroad-sd	-18.61	-38.79	-36.03	-18.43	0.61	1.15	0.97	0.59
	Office-sd	-27.56	-34.13	-35.01	-26.83	0.77	1.30	1.03	0.75
	Overbridge-sd	-37.90	-39.93	-37.48	-37.58	1.77	1.49	1.55	1.71
	Intersection-hd	-18.74	-30.69	-23.81	-18.08	0.71	0.74	0.69	0.66
	Mainroad-hd	-45.65	-28.07	-20.06	-42.29	1.01	0.24	0.23	0.91
	Average	**-30.39**	**-34.35**	**-31.15**	**-29.54**	**1.03**	**1.00**	**0.96**	**0.98**

Table 3 shows the results of the proposed method tested on the low bit-rate setting (on the second group of QPs). When tested on the conference videos, the average BD-BR reductions are 12.14%, 25.32%, and 25.33% on Y, U, and V components respectively, and the corresponding BD-PSNR increments are 0.47dB, 0.70dB, and 0.81dB, respectively. Furthermore, the YUV BD-BR reduction and BD-PSNR increment are 12.46% and 0.47dB. When tested on the surveillance videos, the average BD-BR reductions are 30.39%, 34.35% and 31.15% on Y, U, and V components respectively, and the corresponding BD-PSNR increments are 1.03dB, 1.00dB and 0.96dB respectively. The YUV BD-BR reduction and BD-PSNR increment are 29.54% and 0.48%, respectively. It shows that the average performance gain is larger than the result tested on the first group of QPs. That is, the proposed method is more applicable to the low bit-rate applications.

Fig. 5 shows the R-D curves of the results of 6 sequences. It can be observed that the proposed method can achieve a significant R-D performance improvement in a wide range of bit-rates. Furthermore, for most of the sequence (such as KristenAndSara and FourPeople), the improvement of the low bit-rates is larger than that of the high bit-rates. This is consistent with the above results as shown in Table 3.

As we know, there is an enhanced frame for each optimization period. The enhanced frames are coded with a small QP, which may introduce bit-rate fluctuations. In practice, a small latency is permitted even in the low delay applications, and there is an elementary stream buffer which makes the bit-rate to be smoothed. Taking a latency of 500ms as an example, for the frame rate 30fps (or 60fps), only the average bit number of 15 (or 30) frames should be investigated to evaluate the bit-rate fluctuations.

Fig. 6 shows the bit numbers of the tested frames (Vidyo3 and Overbridge, QP is 32, picture order count: 100 to 300). The black line shows the coded bit number of each frame. The blue line shows the bit numbers smooth with 15 frames, and the red line shows the bit numbers smooth with 30 frames. As it can be seen in this figure, the coded bit numbers of the enhanced frames are larger than that of the other frames. However, after smoothing with 15 or 30 frames, there is no remarkable fluctuation.

TABLE 4. R-D PERFORMANCE IMPROVEMENTS OF THE PROPOSED METHOD COMPARED WITH THE HEVC DEFAULT SCHEME.

Sequences		Proposed Method			[9]			[10]		
		Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)
Conference Videos	FourPeople-720p	-14.50	-35.71	-33.67	-8.02	-15.86	-14.41	-2.75	-7.37	-7.18
	Johnny-720p	-7.22	-36.01	-36.70	1.82	-15.91	-14.53	5.50	-4.05	-4.66
	Kristen&Sara-720p	-12.26	-36.29	-38.44	-9.06	-19.28	-18.70	-3.80	-9.45	-8.40
	Vidyo1-720p	-9.34	-34.46	-36.30	-5.99	-11.15	-13.02	-2.06	-3.61	-5.45
	Vidyo3-720p	-9.43	-40.65	-48.96	-10.10	-16.53	-33.67	-6.47	-9.07	-21.58
	Vidyo4-720p	-11.21	-36.92	-34.86	-0.26	-13.37	-15.18	2.66	-6.02	-6.38
	Average	**-10.66**	**-36.67**	**-38.15**	**-5.27**	**-15.35**	**-18.25**	**-1.15**	**-6.59**	**-8.94**
Surveillance Videos	Bank-sd	-23.68	-53.00	-52.06	-48.88	-72.46	-73.78	-44.79	-66.27	-68.45
	Crossroad-sd	-5.94	-49.97	-48.04	-29.24	-71.06	-67.37	-27.41	-65.50	-61.80
	Office-sd	-15.56	-57.27	-58.54	-16.17	-54.70	-50.88	-17.49	-50.79	-47.22
	Overbridge-sd	-28.07	-55.16	-52.38	-46.91	-71.84	-70.48	-42.96	-65.31	-64.82
	Intersection-hd	-14.27	-27.02	-26.16	-21.45	-33.74	-31.28	-16.58	-27.92	-26.12
	Mainroad-hd	-21.17	-56.39	-46.10	-70.15	-83.13	-75.49	-66.25	-80.06	-71.34
	Average	**-18.11**	**-49.80**	**-47.21**	**-38.80**	**-64.49**	**-61.55**	**-35.91**	**-59.31**	**-56.63**

C. COMPARISON WITH TWO STATE-OF-THE-ART METHODS

Two state-of-the-art methods [9] and [10] are compared with the proposed method. For the fair comparison, the proposed method as well as the compared methods are all implemented on HM 12.0[2]. The test conditions are the same as in [9], in which the frame structure are low delay IBBB and the IBDI depth is set to 10. Table 4 shows the BD-BR saving on Y, U and V components over the anchor (HM 12.0 reference software). As shown in Table 4, the average performance bit-savings of the proposed methods are 10.66%, 36.67% and 38.15% on the conference videos and 18.11%, 49.80% and 47.21% on the surveillance videos. It indicates that the proposed method significantly outperforms the default HEVC scheme for both the conference videos and the surveillance videos.

For the compared methods, the average bit-savings of [9] are 5.27%, 15.35% and 18.25% on the conference videos and 38.80%, 64.49% and 61.55% on the surveillance videos. The average bit-savings of [10] are 1.15%, 6.59% and 8.94% on the conference videos and 35.91%, 59.31% and 56.63% on the surveillance videos. Overall, the proposed method outperforms both the compared methods while coding the conference videos. However, when coding the surveillance videos, the compared methods outperform the proposed method.

In the compared methods, the background frames are generated to reduce the long-term redundance. In this scheme, the performance gains are much smaller on the conference videos than on the surveillance videos. On the conference videos, it is difficult to model a clean background frame, since the backgrounds are usually covered by the moving foregrounds, such as bodies, arms and heads. However, the proposed method is designed based on the decay model, which investigates the variation of the error propagation ratio. The decay model has considered the disturb of the moving foregrounds. Thus, the proposed method is more applicable to coding the conference videos than the compared methods.

Moreover, although the compared methods are more efficient for coding the surveillance videos than the proposed method, the average performance gains of the proposed method are also significant (Y-BDBR: 18.11%, U-BDBR: 49.80%, V-BDBR: 47.21%). This is still meaningful because in the proposed method only the QP offsets are changed and the coding tools are not changed. In particular, the decoder is not changed. This is conducive to the coding technology and product upgrading.

V. 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 is deployed to improve the global performance. The second one is the 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. The proposed method can efficiently exploit the long-term frame dependency of the background regions. Experiments are tested on the surveillance and conference videos. The results show that the proposed algorithm achieves significant performance improvement.

VI. ACKNOWLEDGE

Portions of the research in this paper use the PKU-SVD-A dataset collected under the sponsor of the National Basic Research Program of China and Chinese National Natural

[2] https://hevc.hhi.fraunhofer.de/svn/svn HEVCSoftware/tags/HM-12.0/

Science Foundation.

REFERENCES

[1] J. Gantz and D. D. Reinsel. (2012, "The IDC digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east [online]," https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf, 2012.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[4] K. Ugur and et al., "High performance, low complexity video coding and the emerging HEVC standard," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 12, pp. 1688–1697, Dec. 2010.

[5] W.-J. Han and et al., "Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 12, pp. 1709–1720, Dec. 2010.

[6] J. Xiong, H. Li, Q. Wu, and F. Meng, "A fast HEVC inter CU selection method based on pyramid motion divergence," IEEE Transactions on Multimedia, vol. 16, no. 2, pp. 559–564, 2014.

[7] J. Xiong, H. Li, F. Meng, S. Zhu, Q. Wu, and B. Zeng, "MRF-based fast HEVC inter CU decision with the variance of absolute differences," IEEE Transactions on Multimedia, vol. 16, no. 8, pp. 2141–2153, Dec. 2014.

[8] J. Xiong, H. Li, F. Meng, Q. Wu, and K. N. Ngan, "Fast HEVC inter CU decision based on latent sad estimation," IEEE Transactions on Multimedia, vol. 17, no. 12, pp. 2147–2159, Dec 2015.

[9] X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao, "Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling," IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 4511–4526, Oct 2014.

[10] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding," IEEE Transactions on Image Processing, vol. 23, no. 2, pp. 769–784, Feb 2014.

[11] X. Zhang, Y. Tian, L. Liang, T. Huang, and W. Gao, "Macro-block-level selective background difference coding for surveillance video," in 2012 IEEE International Conference on Multimedia and Expo, July 2012, pp. 1067–1072.

[12] L. Yin, R. Hu, S. Chen, J. Xiao, and J. Hu, "A block-based background model for surveillance video coding," in 2015 Data Compression Conference, April 2015, pp. 476–476.

[13] L. Zhao, X. Zhang, Y. Tian, R. Wang, and T. Huang, "A background proportion adaptive lagrange multiplier selection method for surveillance video on HEVC," in 2013 IEEE International Conference on Multimedia and Expo (ICME), July 2013, pp. 1–6.

[14] X. Li and Y. Abudoulikemu, "Background-foreground information based bit allocation algorithm for surveillance video on high efficiency video coding (HEVC)," in 2016 Visual Communications and Image Processing (VCIP), Nov 2016, pp. 1–4.

[15] S. Zhang, K. Wei, H. Jia, X. Xie, and W. Gao, "An efficient foreground-based surveillance video coding scheme in low bit-rate compression," in 2012 Visual Communications and Image Processing, Nov 2012, pp. 1–6.

[16] C. Chen, J. Cai, W. Lin, and G. Shi, "Incremental low-rank and sparse decomposition for compressing videos captured by fixed cameras," Journal of Visual Communication and Image Representation, vol. 26, pp. 338–348, 2015.

[17] L. Zhao, Y. Tian, and T. Huang, "Background-foreground division based search for motion estimation in surveillance video coding," in 2014 IEEE International Conference on Multimedia and Expo (ICME), July 2014, pp. 1–6.

[18] J. Xiao, R. Hu, L. Liao, Y. Chen, Z. Wang, and Z. Xiong, "Knowledge-based coding of objects for multisource surveillance video data," IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1691–1706, Sept 2016.

[19] C. Ma, D. Liu, X. Peng, and F. Wu, "Surveillance video coding with vehicle library," in 2017 IEEE International Conference on Image Processing (ICIP), Sept 2017, pp. 270–274.

[20] K. Yang, F. Chen, D. Liu, Z. Chen, and W. Li, "Surveillance video coding with dynamic textural background detection," in 2017 IEEE International Conference on Image Processing (ICIP), Sept 2017, pp. 2736–2740.

[21] C. Chen, J. Cai, W. Lin, and G. Shi, "Surveillance video coding via low-rank and sparse decomposition," in Proceedings of the 20th ACM International Conference on Multimedia, ser. MM '12. New York, NY, USA: ACM, 2012, pp. 713–716. [Online]. Available: http://doi.acm.org/10.1145/2393347.2396294

[22] L. Zhao, X. Zhang, X. Zhang, S. Wang, S. Wang, S. Ma, and W. Gao, "Intelligent analysis oriented surveillance video coding," in 2017 IEEE International Conference on Multimedia and Expo (ICME), July 2017, pp. 37–42.

[23] A. Yang, H. Zeng, J. Chen, J. Zhu, and C. Cai, "Perceptual feature guided rate distortion optimization for high efficiency video coding," Multidimensional Systems and Signal Processing, vol. 28, no. 4, pp. 1249–1266, Oct 2017. [Online]. Available: https://doi.org/10.1007/s11045-016-0395-2

[24] Q. Wu, H. Li, F. Meng, K. N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 3, pp. 425–440, March 2016.

[25] Q. Wu, H. Li, Z. Wang, F. Meng, B. Luo, W. Li, and K. N. Ngan, "Blind image quality assessment based on rank-order regularized regression," IEEE Transactions on Multimedia, vol. 19, no. 11, pp. 2490–2504, Nov 2017.

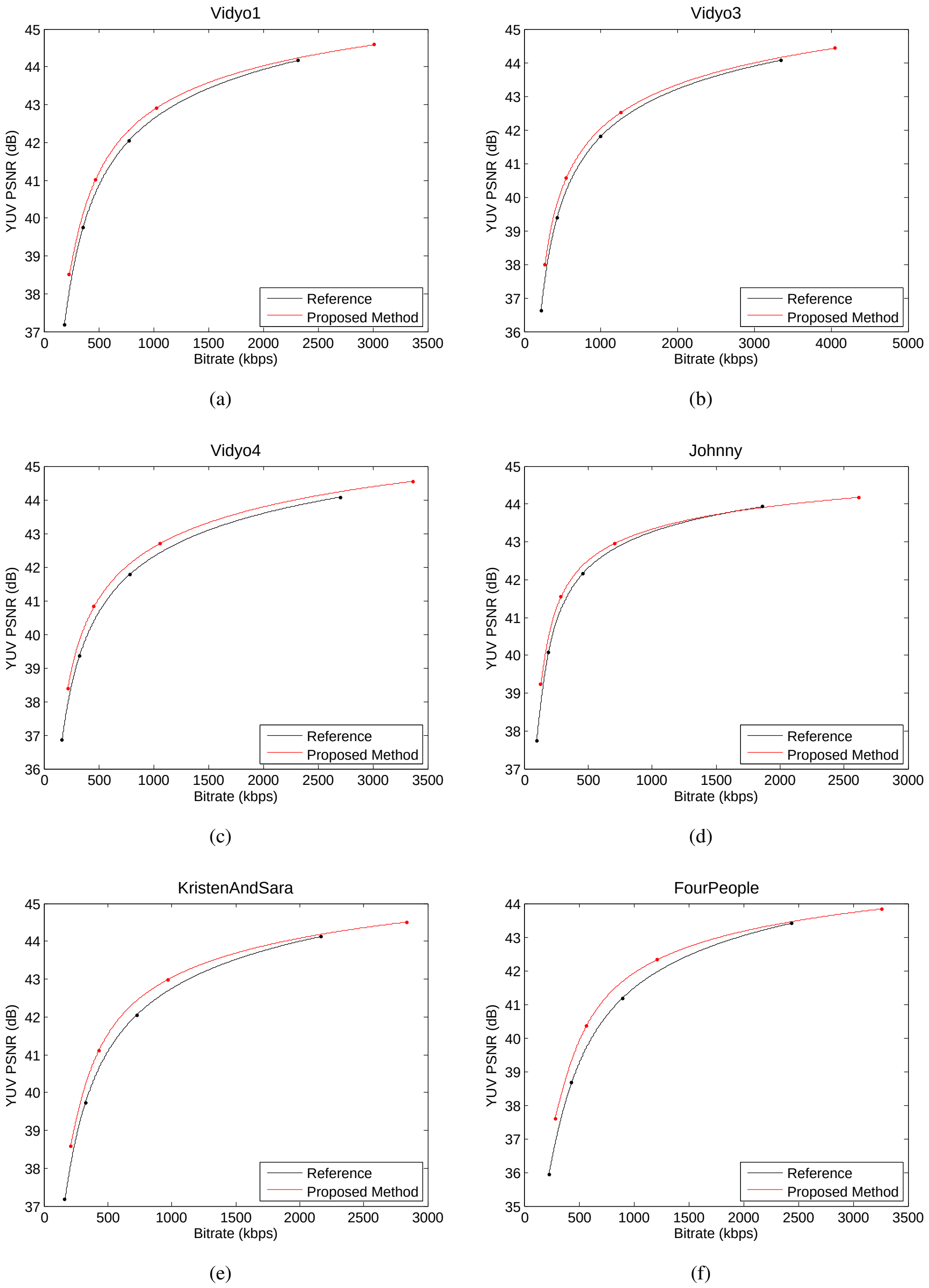
[26] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K. Ma, "Esim: Edge similarity for screen content image quality assessment," IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4818–4831, Oct 2017.

[27] Q. Wu, H. Li, F. Meng, and K. N. Ngan, "A perceptually weighted rank correlation indicator for objective image quality assessment," IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2499–2513, May 2018.

[28] I. K. Kim, J. Min, T. Lee, W. J. Han, and J. Park, "Block partitioning structure in the HEVC standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1697–1706, Dec 2012.

[29] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," no. ITU-T SC16/Q6, VCEG-M33, Austin, USA, April 2001.

[30] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, "The IEEE 1857 standard: Empowering smart video surveillance systems," IEEE Intelligent Systems, vol. 29, no. 5, pp. 30–39, Sept 2014.



(a)

(b)

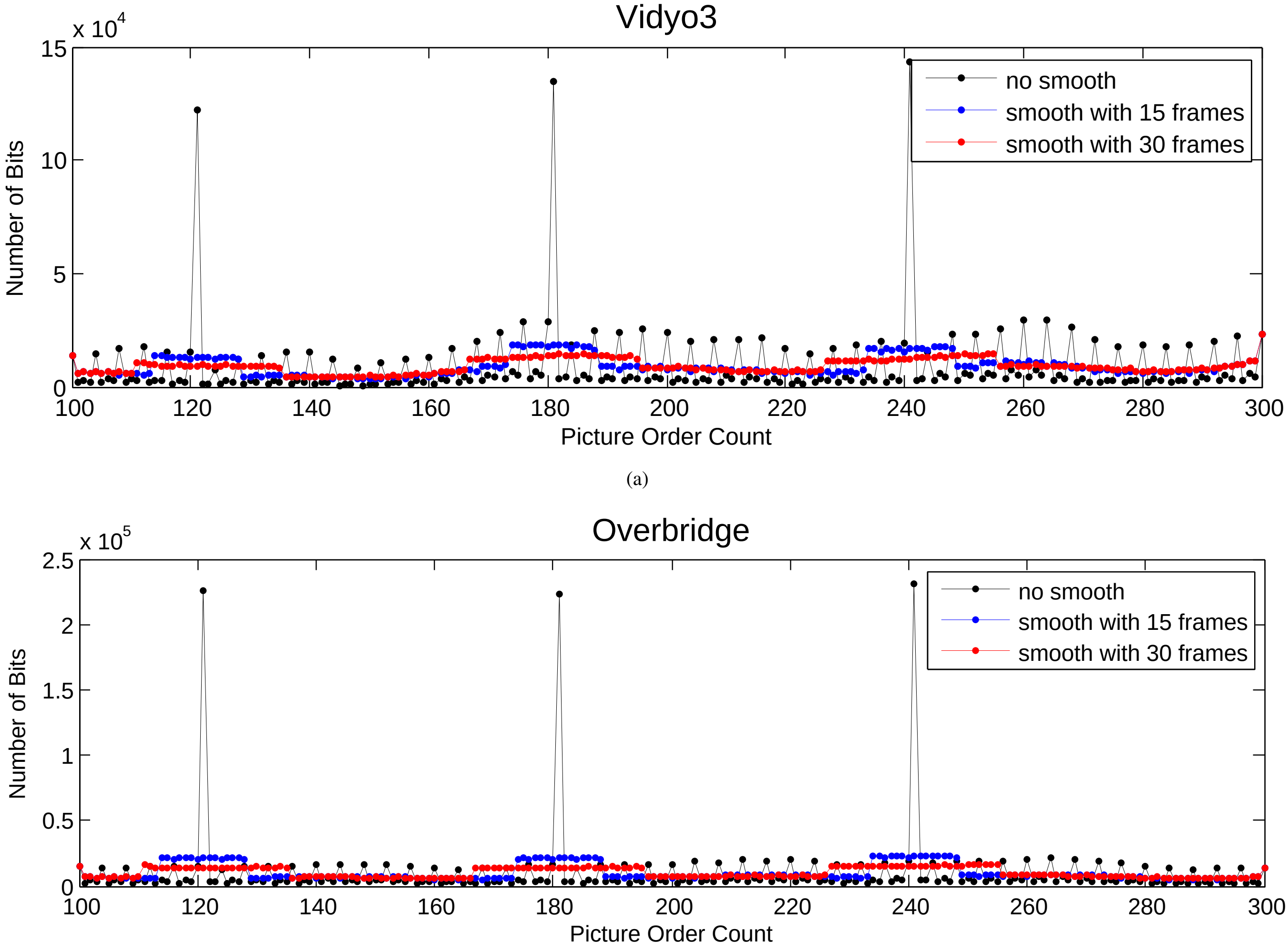
(c)

(d)

(e)

(f)

FIGURE 5. R-D curves of the proposed method compared with the HEVC default scheme. With the sequences: (a) Vidyo1, (b) Vido3, (c) Vidyo4, (d) Johnny, (e) KristenAndSara (f) FourPeople.



(a)

(b)

FIGURE 6. The bit numbers of coded frame (QP is 32). with the sequences: (a) Vidyo3, (b) Overbridge.