

Received March 9, 2019, accepted April 1, 2019, date of publication April 4, 2019, date of current version April 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909270

Periodic Enhanced Frame Based Long-Short-Term Reference in HEVC for Conference and Surveillance Video Coding

JIAN XIONG¹, (Member, IEEE), XIANZHONG LONG², RAN SHI³,
MIAOHUI WANG⁴, (Member, IEEE), QUAN ZHOU¹, (Member, IEEE),
AND GUAN GUI¹, (Senior Member, IEEE)

¹National Engineering Research Center of Communications and Networking, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

²School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

³School of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China

⁴School of Information Engineering, Shenzhen University, Shenzhen 518060, China

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

This work was supported in part by the National Natural Science Foundation of China under Grant 61701258, Grant 61701310, Grant 61801219, and Grant 61876093, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170906 and Grant BK20181393, and in part by the Natural Science Foundation of Jiangsu Higher Education Institutions under Grant 17KJB510044 and Grant 17KJB510038.

ABSTRACT The coding tools of the High Efficiency Video Coding (HEVC) standard were essentially designed for ordinary video contents. They have not fully exploited the long-term temporal dependence of surveillance videos and conference videos, in which the backgrounds are generally stationary. In this paper, a periodic-enhanced frame-based long-short-term reference scheme is presented for coding the videos with stationary backgrounds. First, periodic-enhanced frames are proposed based on global rate-distortion optimization, in which the background error propagation characteristic is introduced to reduce the long-term temporal redundancy. Second, two limitations, such as temporary occlusion and quantization error, are analyzed as applying the enhanced frame on a default HEVC hierarchical prediction structure. In order to reduce the limitations, a long-short-term reference scheme is presented based on the enhanced frames. In this scheme, the periodic-enhanced frames are employed as a long-term reference picture, and the adjacent frames are set as short-term reference pictures. The proposed method is verified by testing on the conference videos and surveillance videos.

INDEX TERMS Surveillance, background, HEVC, reference picture, enhanced frame.

I. INTRODUCTION

Recently, video conference and video surveillance systems are widely applied in the fields of communication and monitor. For instance, video surveillance has rapidly become an important urban infrastructure [1], [2]. As reported in IDC [3], the number of monitor cameras will grow to 5,800 exabytes by 2020. The explosive growth of video data brings a significantly big challenge in transmission and storage.

The latest standard high efficiency video coding (HEVC) [4]–[6] can halve the coding bits when compared with the previous standard as H.264/AVC [7]. That is because, many efficient coding tools have been employed

in HEVC, such as quad-tree based coding units (CU), angular intra prediction, merge modes, sample adaptive offset (SAO) filter and so on. Besides, rate-distortion optimization (RDO) is a key technology, which is adopted to minimize the redundancy and select the optimal coding modes [8]–[13]. However, these coding tools are essentially designed for ordinary video contents. The surveillance and conference videos are special and different from the generic videos [14]. The special characteristics are in two aspects: (1) The cameras in the video surveillance and video conference systems are usually stationary, such that the acquired videos are generally with stationary backgrounds, and these videos have stronger long-term redundancy than the generic videos; (2) The motion patterns are generally simpler than that in the generic video contents, such as rigid moving vehicles in traffic surveillance and small body shaking in conference videos.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan.

These characteristics were inefficiently studied in traditional coding tools.

Related works were presented to exploit the special characteristics of conference and surveillance videos. In [15], the backgrounds and the foregrounds were divided to accelerate the motion estimation (ME) for surveillance video coding. In [16], the authors proposed a bit allocation method based on block classification. In this method, the largest coding units (LCUs) are classified into background CUs and foreground CUs. In [17], a surveillance video coding framework was proposed to improve the intelligent analysis performance. Besides the above works, more concern is given to the compression efficiency for solving the key problem of surveillance and conference video storage and transmission. Recent works were trying to exploit the special characteristics of the stationary background in surveillance and conference videos. These methods are composed of three major classes, including foreground-based schemes [18]–[21], background modeling-based schemes [22]–[25], background-foreground information-based learning schemes [26]–[28] and so on.

Firstly, the foreground-based schemes mainly focus on the processing of foreground objects. In [18], the authors presented a foreground-based coding method for coding surveillance videos at low bit-rate. In this method, the foreground frames were obtained by segmentation, and foreground-based motion compensation was applied to reduce prediction error. In [19], a knowledge-based representation scheme was proposed to mitigate the global redundancy of foreground objects in multisource surveillance videos. In this method, the common knowledge of the objects across multiple cameras, as well as the local variations on the objects' surface, are described as two different parts. In [20], a vehicle library was built to compress surveillance videos. The library was constructed with high quality copies of similar vehicles to reduce the long-term redundancy. In [21], the authors presented a dynamic texture coding method to compress surveillance videos. In this method, the dynamic textural contents are detected and skipped in the encoder and restored by using dynamic texture synthesis. The foreground-based methods mainly work on the applications of specific foreground objects, such as vehicles, dynamic textures. However, compared with the foreground regions, there is more long-term redundancy in background regions of the conference and surveillance videos.

Secondly, the background modeling-based schemes generally construct background frames to mitigate the long-term redundancy. In [22], a background difference coding method was proposed by selecting the Macro-blocks to be encoded as one of the two manners, which include encoding the original MB and encoding the MB difference. In [23], an adaptive prediction method was presented on basis of background modeling. In this method, the long-term redundancy of background regions was mitigated by conducting prediction on the generated background pictures. In [24], the authors presented a background modeling-based prediction method in order to improve the performance of prediction. The background

pictures were generated from train sets of the original pictures and encoded as intra frames with smaller quantization parameter (QP). The background pictures are employed as long-term reference frames in the hierarchical prediction structure. Nevertheless, frame-based background modeling cannot generate the backgrounds efficiently, in which the background may become worse as the frame index increases. In order to mitigate the issue, the background is modeled based on the block level as in [25]. In this method, backgrounds were generated on basis of coding units (CUs). Furthermore, the backgrounds are updated for every frame that is different from being updated for each GOP. However, the block-based methods may suffer from block artifacts between the foregrounds and backgrounds.

Finally, the third class of schemes is the learning-based methods which employ the foreground-background information. In [26], [27], the strong temporal redundancy of surveillance videos was reduced by sparse and low-rank decomposition. In [28], the Lagrangian multiplier was adaptively selected for coding the videos with stationary backgrounds. In this method, the lambda factor parameters are learned to change the Lagrangian multiplier in RDO. The feature of background proportions was employed in the training process. However, the lambda factor parameters are limited for fully reducing the long-term redundancy. In addition, there is extra computation in the training process.

The background modeling based methods can significantly improve the coding performance of surveillance videos, but they are not significantly efficient in coding conference videos. The reason is that, compared with the surveillance videos, the foreground regions of the conference videos are generally large and the motion of foregrounds is more intensive. The motion of the foregrounds is bad for generating a clean background. In this paper, a new prediction structure is proposed based on a periodic enhanced frame (PEF). On the basis of HEVC, the proposed method try to improve the coding efficiency of not only the surveillance videos but also the conference video. Firstly, a line model is presented to describe the error propagation characteristic of backgrounds. In this model, the temporal dependency between the frames is investigated as the error propagation ratio. Based on this model, we present an enhanced frame-based RDO scheme, in which the enhanced frames are encoded with small QPs. Secondly, we analyze the limitations of the default HEVC hierarchical prediction structure for applying the enhanced frame-based RDO. In order to mitigate the limitations, we propose a PEF-based long-short-term reference scheme. In this scheme, the enhanced frames are periodically encoded and employed as the long-term references, in which not only the backgrounds but also the foregrounds are periodically improved in the enhanced frames. The results show that the proposed method can significantly outperform the background modeling based methods for coding the conference videos. Furthermore, it can achieve competitive results as that of the background modeling methods when coding the surveillance videos.

The remainder of the paper is organized as follows. A brief description of HEVC default coding tools is presented in Section II. The enhanced frame-based RDO method is proposed in Section III. The PEF-based long-short-term reference scheme is given in Section IV. Experiments and more discussions are provided in Section V. In the end, we make some conclusion in Section VI.

II. MOTION COMPENSATION AND MOTIVATION

A. MOTION COMPENSATION AND HIERARCHICAL PREDICTION STRUCTURE IN HEVC

Motion compensation (MC) and motion estimation (ME) are the fundamental techniques used in existing video coding standards including H.264/AVC and HEVC. In order to reduce the temporal redundancy, MC is employed to predict a block in a frame based on its previous frames in the decoding order, named the reference frames. In this method, the predicted block is obtained from the reference frames with a motion vector (MV) via ME. The differences between the predicted block and the original block are called as residual errors, which are given by

$$e(f, x) = I(f, x) - I'(f - \Delta f, x + v), \quad (1)$$

where $I(f, x)$ denotes pixel value at position x in frame f , and $I'(f - \Delta f, x + v)$ denotes the reconstructed pixel at position $x + v$ in the reference frame denoted as $f - \Delta f$. The symbols v and $e(f, x)$ denote the MV and the corresponding residual signal obtained by ME and MC, respectively. Subsequently, only the residual errors need to be encoded. It should be noticed that a better reference can improve the coding efficiency.

In the MC method, the recent frames are generally employed as the candidate reference frames, since they have strong redundancy with the current frame. The low-delay setting in HM (HEVC test model) is a good option for the real-time coding of surveillance videos and conference videos [24]. Take the low-delay setting as an example, in this setting, a hierarchical prediction structure (HPS) is utilized to select the reference frame. As shown in Figure. 1, the intra encoded frame is shown as the yellow bar, and the inter encoded frames are shown as the red bars. Furthermore, the corresponding QP offsets are shown as the numbers in the gray box below. The frames in each group of pictures (GOP) are hierarchically quantized and the last frame in each GOP is quantized with a smaller QP. The last frame in each GOP

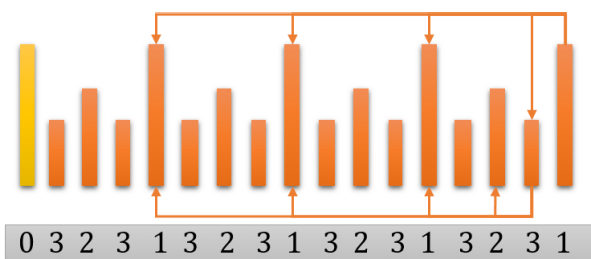


FIGURE 1. The HEVC default hierarchical prediction structure (HPS).

is named as a more important frame (MIF). The reference picture list is composed of its previous frame and 3 previous MIFs. The reference structure plays an important role in the MC prediction. That is, an efficient reference scheme will result in improved coding performance.

B. PROBLEM STATEMENT

In order to improve the efficiency of MC, background modeling based methods [23], [24] are proposed for coding the surveillance videos. Since a huge number of static background regions widely exist in the surveillance videos, some of the surveillance videos can be considered as quasi-static videos, i.e., most of the regions are static. In that case, the redundancy of surveillance videos can be further reduced by using background modeling.

However, for conference videos, background modeling is not a significantly efficient scheme [23]. Experiments are tested to analyze the reasons. In the experiments, mean absolute errors (MAEs) between the frames are used to describe the motion amplitudes of the foreground objects. Figure 2 shows the MAEs with different frame intervals of surveillance videos (such as Bank, Overbridge and Mainroad) and conference videos (such as Vidyo1, Vidyo3 and Johnny). Firstly, it is obvious that the MAEs of the conference videos are larger than that of the surveillance videos. Secondly, it also can be found that, as the frame interval increases, the MAEs of the conference videos also increase. However, the MAEs of the surveillance videos vary on small scales. To sum up, compared with the surveillance videos, this conference videos have the following characteristics: (1) the foreground objects (usually persons) of conference videos are more closer to the camera, such that the proportions of static backgrounds are related small; (2) the motion amplitudes of the foreground objects are related small, and the foreground objects remain stationary for relatively short periods, but are moved in long terms. These features make the background modeling based methods are not significantly efficient in coding conference videos.

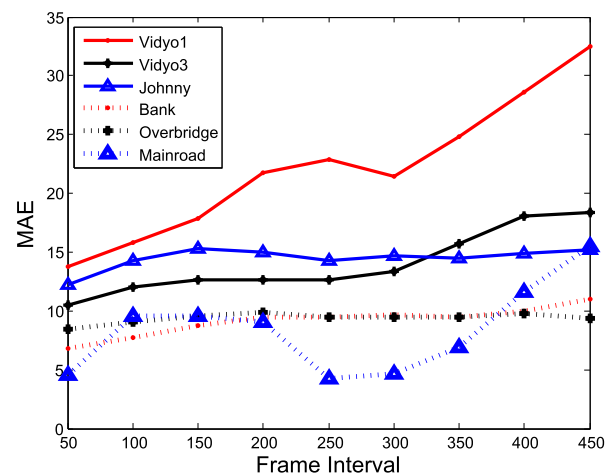


FIGURE 2. Mean absolute errors (MAEs) between the frames with different intervals.

Therefore, aiming at the characteristics of conference videos, we propose a periodic enhanced frame based long-short-term reference coding scheme. In this scheme, not only the backgrounds but also the foregrounds are periodically improved in the enhanced frames. The periodic enhanced scheme makes it is possible to further reduce the redundancy of the foregrounds, which remain stationary for short periods.

III. ENHANCED FRAME-BASED RATE DISTORTION OPTIMIZATION

A. RATE DISTORTION OPTIMIZATION

RDO technology plays an important role in existing video coding methods. In the methods, the optimal coding mode and parameters are selected by using RDO, which minimize the total distortion under a constraint of coding bits. The constraint problem can be translated into an unconstrained problem of minimizing the Rate-Distortion cost, which is given by

$$\min(J = \sum_{f=1}^k D_f + \lambda \cdot \sum_{f=1}^k R_f). \quad (2)$$

where the symbols R_f and D_f denote the bit consumption and the distortion of frame f , ($f = 1, 2, \dots, k$), respectively. The symbol k denotes the total number of the coded frames. The parameter λ denotes the Lagrangian multiplier which is introduced to balance the distortion and bit consumption, which is computed as a function of QP value. In the default HPS, the QP offsets are the same between the GOPs, such that the Lagrangian multipliers are the same between the GOPs.

B. BACKGROUND ERROR PROPAGATION

Since the surveillance videos and conference videos are usually acquired by stationary cameras, there are a huge amount of background pixels. The background pixels are generally stationary in a long time, i.e.,

$$I(k, x) = I(k-1, x) = \dots = I(1, x). \quad (3)$$

Accordingly, without considering the local illumination variation, the background blocks are generally encoded as SKIP mode with zero MV, in which there is no residual signal to be encoded. That is, the co-located reconstructed background pixels of the consecutive frames are the same, i.e.,

$$I'(k, x) \approx I'(k-1, x) \approx \dots \approx I'(1, x). \quad (4)$$

In that case, the MC difference is equal to the reconstruct distortion. It can be concluded that, the distortion of co-located background CUs are approximately the same in the consecutive frames, i.e.,

$$SSD(k, C) \approx SSD(k-1, C) \approx \dots \approx SSD(1, C), \quad (5)$$

where $SSD(k, C)$ denotes the sum of squared difference between the original values and the reconstructed values of CU C in frame k , i.e., $SSD(k, C) = \sum_{x \in C} (I(k, x) - I'(k, x))^2$ (distortion of CU C). This is the so-called error propagation

characteristic of backgrounds. As shown in Figure 3, the coding errors in the stationary backgrounds propagate from the first frame to its subsequent frames successively.

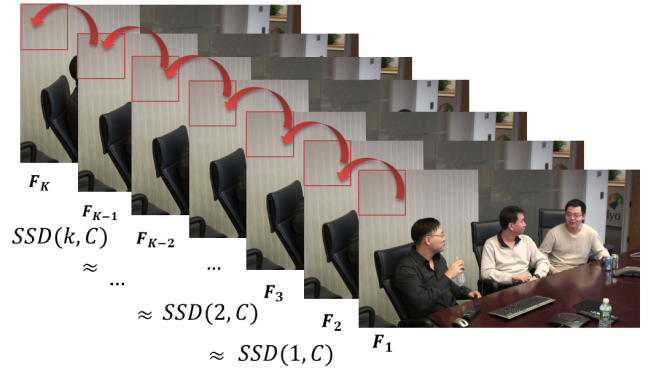


FIGURE 3. The characteristic of background error propagation.

For each of the subsequent frames, the total distortion can be divided into two parts: the background distortion being propagated from the first frame (denoted as D_{pro}), and the distortion of the non-propagated regions (denoted as D_{non}). That is, the distortion of frame f can be expressed as

$$D_f = D_{pro}(f) + D_{non}(f). \quad (6)$$

Since $D_{pro}(f)$ is propagated from the first frame, it can be described as a part of its total distortion, as

$$D_{pro}(f) = r_f \cdot D_1. \quad (7)$$

The parameter r_f , ($0 < r_f < 1$) is the propagation ratio, which is related to the temporary dependency between the first frame and frame f (always related to the background proportion). Thereby, we can obtain,

$$D_f = r_f \cdot D_1 + D_{non}(f). \quad (8)$$

The linear model shows that the distortion of the subsequent frames have a strong linear relationship with that of the first frame.

C. ENHANCED FRAME-BASED RDO

Based on the linear model, (2) can be changed as,

$$\min(D_1 \cdot \sum_{f=1}^k r_f + \sum_{f=1}^k D_{non}(f) + \lambda R_1 + \lambda \cdot \sum_{f=2}^k R_f). \quad (9)$$

The minimal R-D cost can be obtained by setting the derivative to be zero. Thus, the Lagrangian parameter of the first frame (denoted as λ_1) is solved as,

$$\lambda_1 = -\frac{\partial D_1}{\partial R_1} = \frac{\lambda}{\sum_{f=1}^k r_f}. \quad (10)$$

In the default HEVC coding scheme, the Lagrangian parameters are computed from QP values. Accordingly, a small QP can be derived from λ_1 as $qp' = qp - 3 \log_2(\sum_{f=1}^k r_f)$, where qp denotes the original QP value

corresponding to λ . That is, the first frame is encoded with the small QP, such that the coding efficiency of the first frame can be improved. We call the frame as the enhanced frame. It can be utilized to lift the compression efficiency of the subsequent frames, and thereby the total coding efficiency.

IV. PERIODIC ENHANCED FRAME-BASED HIERARCHICAL PREDICTION

A. LIMITATIONS OF ENHANCED FRAME-BASED RDO IN DEFAULT HPS

As declared in the above section, an enhanced frame can be utilized to lift the compression efficiency of the total frames, by exploiting the background error propagation characteristic. However, simply using the enhanced frame in the HEVC default HPS cannot take full advantage of the characteristic. There are two main limitations, including the temporary occlusion and quantization error.

(1) Temporary occlusion

The main idea of the enhanced frame-based RDO is exploiting the characteristic of background error propagation. In this scheme, the errors should propagate to the subsequent frames successively. However, in the default HEVC HPS, only the backgrounds which satisfy the propagation characteristic can benefit from the scheme. This is due to the default reference structure, in which, only the recent frames (the previous frame and 3 previous MIFs) are employed in the reference picture list. The propagation characteristic will be undermined by temporary occlusion or local illumination variation, which is also frequently appearing in surveillance and conference videos. As the frame index increases, more and more backgrounds are occluded by the motion foreground objects, i.e., the impact of the first (enhanced) frame on the subsequent frames gradually decreases frame by frame.

(2) Quantization error

As we know, a better reference in MC can improve the coding efficiency. In the HEVC default HPS, the recent frames are utilized as the reference frames. That is, the subsequent frames are predicted on their previous frames successively. When conducting ME on the stationary backgrounds of conference and surveillance videos, it is reasonable to assume the MVs to be zero. Combining (1) and (3), the MC difference of pixel $I(f, x)$ can be written as,

$$\begin{aligned} e(f, x) &= I(f, x) - I'(f - \Delta f, x) \\ &= I(f - \Delta f, x) - I'(f - \Delta f, x) \end{aligned} \quad (11)$$

where $f - \Delta f$ denotes the reference frame, and $I'(f - \Delta f, x)$ is the corresponding reference pixel. The symbol $e(f, x)$ is given as the MC difference. The above equation shows that the MC difference is equal to the reconstructed error of the reference pixel. The reconstructed errors are mainly affected by the QP values. That is, the effect of MC is mainly influenced by the QP values of the reference frames. However, except for the enhanced frame, the QP values of the other frames are not changed in the default HEVC HPS.

B. ENHANCED FRAME-BASED LONG-SHORT-TERM REFERENCE SCHEME

The above limitations can be mitigated by making the enhanced frame to be a long-term reference. Specifically, in the reference picture list, the fourth reference frame is replaced by the enhanced frame. That is, the reference picture list is composed of three short-term reference pictures plus another one long-term reference picture. We named it as the enhanced frame-based long-short-term reference scheme. The proposed scheme can maximize the effect of the enhanced frame.

On one hand, the long-term reference scheme will not be affected by the temporary occlusion and local illumination variation. In this scheme, as long as the same background CU appear in both the enhanced frame and one subsequent frame, MC of the background CU may be directly carried out on the enhanced frame, such that the performance gains of the enhanced frame are not necessarily to be propagated frame by frame.

On the other hand, the enhanced frame-based long-term reference will lead to better reference. In this scheme, MC of background CUs is directly conducted on the enhanced frame. Compared with coding the subsequent frames, smaller QP is applied to code the enhanced frames, such that the quantization errors are significantly smaller than that of the subsequent frames. In that case, combining (1) and (3), the MC difference can be written as,

$$\begin{aligned} e(f, x) &= I(f, x) - I'(1, x) \\ &= I(1, x) - I'(1, x) \end{aligned} \quad (12)$$

It should be noted that the CUs in the first frame are encoded with a small QP, such that the distortion are smaller than the other frames.

C. OBSERVATION AND ANALYSIS

In the above subsections, a new prediction structure is presented based on the enhanced frame. Firstly, an enhanced frame-based RDO is proposed to exploit the temporal dependency between the enhanced frame and its subsequent frames. Secondly, based on the enhanced frame, a long-short-term reference scheme is presented by employing the enhanced frame as a long-term reference. The main ideas of both the schemes are to make maximum use of the temporal dependency between the enhanced frame and the subsequent frames.

Experiments are carried out to investigate the variation of temporal dependency, i.e., r_f in (8). In the experiments, three groups of tests are carried out on the first 200 frames of two video sequences (Vidyo1 and Bank). In these tests, the initial QP value is set to 32. The first group is the anchor, which is conducted on the default HPS. The second group is also performed on the default HPS. However, the first inter frame (enhanced frame) is encoded with a QP offset as -12 , and the other frames are encoded with the same QP values as in the anchor. That is, the second group conducts the enhanced

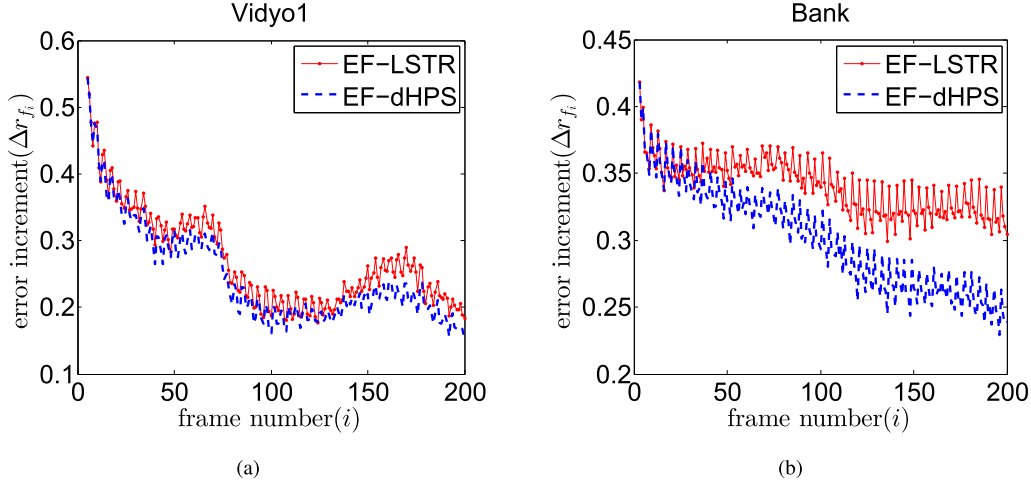


FIGURE 4. The variation of error propagation ratios. With the sequences: (a) Vidyo1, (b) Bank.

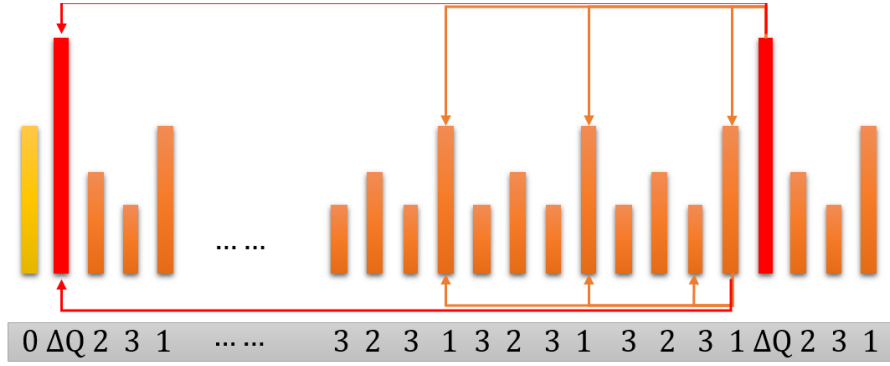


FIGURE 5. The proposed periodic enhanced frame based long-short-term reference (PEF-LSTR) scheme.

frame-based RDO on the default HPS, denoted as EF-dHPS. In the third group, the QP values are the same as the second group, i.e., there is also an enhanced frame. Nevertheless, the enhanced frame is employed to replace the fourth reference picture. The third group is denoted as EF-LSTR.

Since r_f is the slope of line model, it can be calculated as the ratio between distortion increments, which is given by,

$$r_f = \frac{\Delta D_f}{\Delta D_1} \quad (13)$$

The distortion increments are calculated by comparing with the anchor. The results are shown in Figure 4. The dark lines show the results of EF-dHPS. The red lines show the results of EF-LSTR. It can be observed that most of the values are higher than 0.2 in both the dark lines and the red lines. It indicates that the enhanced frame can exploit the temporal dependency efficiently. Furthermore, the red lines are higher than the dark lines, i.e., making the enhanced frame to be the long term reference can increase the temporal dependency.

However, for both of the two groups, the value of r_f has trended down as the frame index increases. It indicates that the temporal dependency also decreases even employing the

enhanced frame as the long-term reference. Therefore, it is necessary to set the enhanced frames periodically. It is called as periodic enhanced frame-based long-short-term reference scheme, denoted as PEF-LSTR. The period length can be set by experiment (discussed in section V-C).

D. OVERALL FRAMEWORK

The PEF-LSTR scheme is shown in Figure 5. In this figure, the bars with different colors are shown as the frames encoded with different modes. The intra-coded frame is shown as the yellow bar. Beyond that, the other bars are the inter-coded frames, in which the red bars are shown as the enhanced frames. The corresponding QP offsets are shown as the numbers in the gray box below. It can be observed that a small QP (minus a large QP offset ΔQ) is employed to encode the enhanced frames. The subsequent frames are encoded with the default QPs (as in the default HPS). In coding each of the subsequent frames, its previous frame and two previous MIFs are employed as the short-term reference pictures. Moreover, the enhanced frame is utilized as the fourth reference picture, i.e., the long-term reference picture. It should be noticed that, because of the moving foreground, the temporal dependency

TABLE 1. Overall experimental results of the proposed method on the anchor.

Sequences		BD-BR (%)			BD-PSNR (dB)		
		Y (%)	U (%)	V (%)	Y (dB)	U (dB)	V (dB)
Surveillance Videos	Bank-sd	-35.22	-64.69	-65.30	0.57	2.00	2.52
	Crossroad-sd	-10.50	-60.95	-57.88	0.41	1.93	1.54
	Office-sd	-20.29	-67.63	-67.80	0.37	2.20	1.68
	Overbridge-sd	-36.21	-64.79	-62.90	1.05	2.74	2.91
	Intersection-hd	-25.27	-31.21	-31.22	0.89	0.67	0.70
	Mainroad-hd	-41.70	-46.73	-41.47	0.92	0.36	0.36
	Average	-28.20	-56.00	-54.43	0.70	1.65	1.62
Conference Videos	FourPeople-720p	-19.18	-38.44	-37.01	0.70	1.15	1.13
	Johnny-720p	-20.12	-44.39	-45.46	0.47	1.15	1.19
	KristenAndSara-720p	-17.89	-39.72	-42.27	0.55	1.07	1.12
	Vidyo1-720p	-15.40	-36.78	-39.14	0.50	0.71	0.95
	Vidyo3-720p	-18.74	-43.38	-53.07	0.60	0.89	1.74
	Vidyo4-720p	-18.68	-41.25	-40.96	0.56	1.01	1.09
	Average	-18.34	-40.66	-42.98	0.56	1.00	1.20

will decrease as the frame index increases. Therefore, in order to exploit the temporal dependency efficiently, the enhanced frames are adopted in a periodic manner in the proposed method.

V. EXPERIMENTAL RESULTS

A. EXPERIMENT SETTINGS

In this section, the proposed PEF-LSTR method, as well as that of the state-of-the-art works [23] and [24], are evaluated by comparing with the default HEVC reference software. These methods are evaluated in terms of the Bjontegaard Delta Peak Signal to Noise Ratio (BD-PSNR) and Bjontegaard Delta bit-rate (BD-BR) [29]. To be fair, all the methods are implemented on the reference software, HM12.0.²

The low delay main setting with the IBDI depth set as 10 (“encoder_lowdelay_main10.cfg”) is adopted in the experiments. In this setting, the frame structure is IBBB, and the GOP size is 4. Coding tools such as fast ME search, RDO quantization, and SAO filter are enabled. QP values including 22, 27, 32 and 37 are tested to cover a wide range of bit-rates. The QP offset of the enhanced frames is set to -12. The period lengths are set to 120 and 200 for conference videos and surveillance videos respectively.

In order to evaluate the proposed method, 12 video sequences with stationary backgrounds are tested in the experiments. The tested video sequences are composed of 6 surveillance videos and 6 conference videos. As shown in Table 1, the conference videos include Vidyo1, Vidyo3, Vidyo4, FourPeople, Kristen&Sara, and Johnny. The surveillance video sequences include Bank, Office, Intersection, Crossroad, Overbridge, and Mainroad. These videos are obtained from the dataset, named PKU-SVD-A [23], [24], [30]. The resolutions of the tested videos include 720 × 576 (SD), 1280 × 720 (720p), and 1600 × 1200 (HD).

B. OVERALL IMPROVEMENT OF THE PROPOSED METHOD

The proposed PEF-LSTR method is compared with the HEVC default method (as the anchor). The results are all displayed in Table 1. It shows that the proposed algorithm achieves significant improvement on the anchor. When encoding the conference videos, the average BD-BR reductions on the Y component is 18.34%, and the average BD-PSNR increments of the Y component is 0.56dB. Moreover, the improvement on U and V components is more significant than that of Y component. Specifically, the average BD-BR reductions on U and V components are 40.66% and 42.98%, and the average BD-PSNR increments on U and V components are 1.00dB and 1.20dB, respectively. When encoding the surveillance videos, the average BD-BR reductions on the Y component is 28.20%, and the average BD-PSNR increments of the Y component is 0.70dB. Furthermore, the average BD-BR reductions on U and V components are 56.00% and 54.43%, and the average BD-PSNR increments on U and V components are 1.65dB and 1.62dB, respectively. It indicates that, in both cases of coding the surveillance videos and the conference videos, the proposed PEF-LSTR method can significantly exceed the HEVC default method.

It also can be observed that the average improvement of coding the surveillance videos is better than that of coding the conference videos. That is because there are a larger proportion of stationary backgrounds in the surveillance videos than in the conference videos. i.e., more background CUs can benefit from the motion compensation on the enhanced frames. Moreover, the results are the same even in the surveillance videos. For instance, the sequences (such as Bank, Overbridge, and Mainroad) have a larger proportion of stationary backgrounds than that of the sequences such as Office and Crossroad. Accordingly, the performance gains of the former sequences are also larger than that of the latter sequences. It is reasonable to conclude that the proposed PEF-LSTR scheme is more suitable to the videos with larger proportions of stationary backgrounds. In other words, it can take full advantage of the background error propagation characteristic.

²https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-12.0/

TABLE 2. Experimental results of the proposed method tested with different period lengths.

Sequences		P = 60			P = 120			P = 200		
		Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)
Surveillance Videos	Bank-sd	-31.33	-58.28	-58.47	-34.46	-63.16	-63.27	-35.22	-64.69	-65.30
	Crossroad-sd	-7.58	-59.61	-57.64	-9.92	-60.96	-58.55	-10.50	-60.95	-57.88
	Office-sd	-18.30	-63.90	-64.84	-20.95	-67.08	-67.76	-20.29	-67.63	-67.80
	Overbridge-sd	-32.26	-58.68	-56.83	-34.63	-62.75	-60.95	-36.21	-64.79	-62.90
	Intersection-hd	-20.78	-32.08	-31.68	-21.74	-31.21	-31.22	-25.27	-33.83	-33.76
	Mainroad-hd	-31.69	-53.54	-46.59	-38.18	-48.66	-41.74	-41.70	-46.73	-41.47
	Average	-23.65	-54.35	-52.68	-26.65	-55.64	-53.92	-28.20	-56.44	-54.85
Conference Videos	FourPeople-720p	-17.04	-37.12	-35.37	-19.18	-38.44	-37.01	-19.00	-37.66	-36.10
	Johnny-720p	-14.69	-39.89	-41.30	-20.12	-44.39	-45.46	-16.46	-44.44	-45.30
	KristenAndSara-720p	-13.85	-37.48	-39.89	-17.89	-39.72	-42.27	-19.96	-41.49	-43.42
	Vidyo1-720p	-12.38	-36.61	-38.54	-15.40	-36.78	-39.14	-14.46	-34.25	-36.77
	Vidyo3-720p	-15.67	-43.82	-50.95	-18.74	-43.38	-53.07	-16.84	-40.17	-52.83
	Vidyo4-720p	-17.12	-40.33	-38.57	-18.68	-41.25	-40.96	-19.19	-40.34	-41.41
	Average	-15.12	-39.21	-40.77	-18.34	-40.66	-42.98	-17.65	-39.72	-42.64

TABLE 3. Experimental results of the proposed two schemes (PEF-dHPS and PEF-LSTR) (BD-BR(%)).

Sequences		PEF-dHPS			PEF-LSTR		
		Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)
Surveillance Videos	Bank-sd	-23.68	-53.00	-52.06	-35.22	-64.69	-65.30
	Crossroad-sd	-5.94	-49.97	-48.04	-10.50	-60.95	-57.88
	Office-sd	-15.56	-57.27	-58.54	-20.29	-67.63	-67.80
	Overbridge-sd	-28.07	-55.16	-52.38	-36.21	-64.79	-62.90
	Intersection-hd	-14.27	-27.02	-26.16	-25.27	-31.21	-31.22
	Mainroad-hd	-21.17	-56.39	-46.10	-41.70	-46.73	-41.47
	Average	-18.11	-49.80	-47.21	-28.20	-56.00	-54.43
Conference Videos	FourPeople-720p	-14.50	-35.71	-33.67	-19.18	-38.44	-37.01
	Johnny-720p	-7.22	-36.01	-36.70	-20.12	-44.39	-45.46
	KristenAndSara-720p	-12.26	-36.29	-38.44	-17.89	-39.72	-42.27
	Vidyo1-720p	-9.34	-34.46	-36.30	-15.40	-36.78	-39.14
	Vidyo3-720p	-9.43	-40.65	-48.96	-18.74	-43.38	-53.07
	Vidyo4-720p	-11.21	-36.92	-34.86	-18.68	-41.25	-40.96
	Average	-10.66	-36.67	-38.15	-18.34	-40.66	-42.98

C. PERIOD LENGTH SETTING

The enhanced frames are adopted in the proposed PEF-LSTR method periodically. Different period lengths may bring about distinct performance. In order to study that, 3 different lengths including 60, 120 and 200, are tested in the experiments. The results are displayed in Table 2. It shows that when we extend the period length from 60 to 120, the average BD-BR reductions of Y component have increased 3% for both the surveillance videos and the conference videos. Since the enhanced frames are encoded with smaller QP values, the encoded bits are significantly higher than that of the other frames. A bigger period length will lead to less number of enhanced frames, thereby, fewer bits will be used to encode the enhanced frames.

However, the period length cannot be extended without limit. When we extend the period length from 120 to 200, the average BD-BR reduction of the Y component has increased only 1.55% for the surveillance videos. What is more, it has decreased by 0.69% for the conference videos. The major reason is that, as the frame order number goes by, the moving foregrounds make the temporal dependency between the subsequent frame and the enhanced frame to be small. In that case, a new enhanced frame is needed.

Specifically, in the proposed method, the period length is set to 120 in the case of coding the conference videos. However, since there are larger proportions of backgrounds in the surveillance videos than that of the conference videos, the period length is set to 200 in the case of coding the surveillance videos.

D. IMPROVEMENT OF DIFFERENT PARTS

As presented in the above sections, the proposed method is composed of two parts: a periodic enhanced frame (PEF) based RDO and PEF-based long-short-term reference. Two groups of experimental results are compared to investigate the contributions of each part. Firstly, the PEF-based RDO is conducted with the default HEVC HPS (denoted as PEF-dHPS). Secondly, the PEF-based RDO is conducted with the long-short-term reference (i.e., PEF-LSTR). The results are shown in Table 3.

When the PEF-based RDO is conducted on the default HPS, the average BD-BR reductions of the Y component are 10.66% and 18.11% for the conference videos and the surveillance videos, respectively. It indicates that the PEF-based RDO achieves a better coding performance than the HEVC reference software. The PEF-based RDO can take

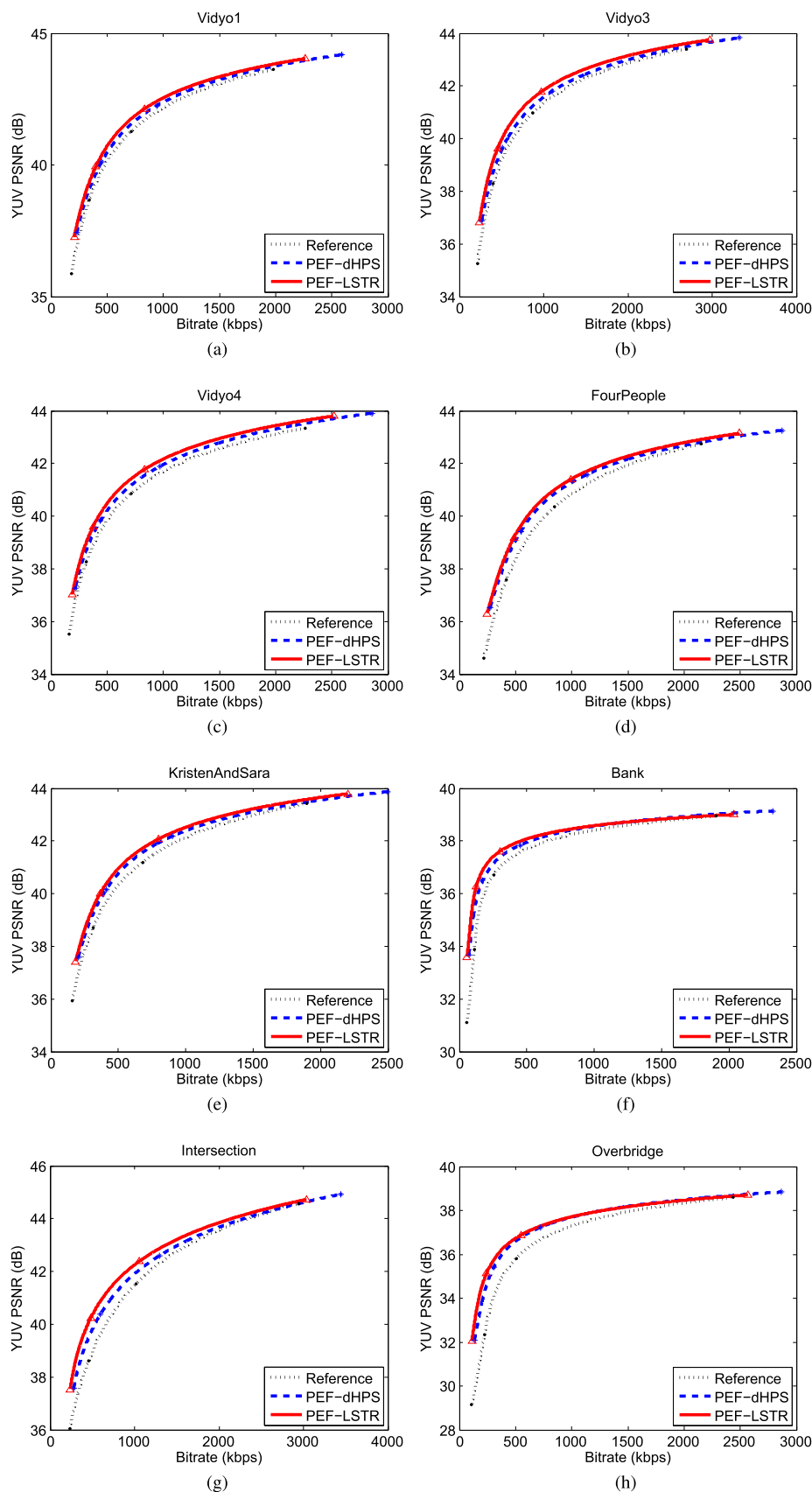


FIGURE 6. R-D curves of the proposed schemes (PEF-dHPS and PEF-LSTR) compared with the HEVC anchor.

TABLE 4. Experimental results compared with the state-of-the-art methods.

Sequences		Proposed Method			[24]			[23]		
		Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)
Surveillance Videos	Bank-sd	-35.22	-64.69	-65.30	-48.88	-72.46	-73.78	-44.79	-66.27	-68.45
	Crossroad-sd	-10.50	-60.95	-57.88	-29.24	-71.06	-67.37	-27.41	-65.50	-61.80
	Office-sd	-20.29	-67.63	-67.80	-16.17	-54.70	-50.88	-17.49	-50.79	-47.22
	Overbridge-sd	-36.21	-64.79	-62.90	-46.91	-71.84	-70.48	-42.96	-65.31	-64.82
	Intersection-hd	-25.27	-33.83	-33.76	-21.45	-33.74	-31.28	-16.58	-27.92	-26.12
	Mainroad-hd	-41.70	-46.73	-41.47	-70.15	-83.13	-75.49	-66.25	-80.06	-71.34
Average		-28.20	-56.44	-54.85	-38.80	-64.49	-61.55	-35.91	-59.31	-56.63
Conference Videos	FourPeople-720p	-19.18	-38.44	-37.01	-8.02	-15.86	-14.41	-2.75	-7.37	-7.18
	Johnny-720p	-20.12	-44.39	-45.46	1.82	-15.91	-14.53	5.50	-4.05	-4.66
	KristenAndSara-720p	-17.89	-39.72	-42.27	-9.06	-19.28	-18.70	-3.80	-9.45	-8.40
	Vidyo1-720p	-15.40	-36.78	-39.14	-5.99	-11.15	-13.02	-2.06	-3.61	-5.45
	Vidyo3-720p	-18.74	-43.38	-53.07	-10.10	-16.53	-33.67	-6.47	-9.07	-21.58
	Vidyo4-720p	-18.68	-41.25	-40.96	-0.26	-13.37	-15.18	2.66	-6.02	-6.38
Average		-18.34	-40.66	-42.98	-5.27	-15.35	-18.25	-1.15	-6.59	-8.94

advantage of the background error propagation characteristic to some extent. However, when the PEF-based RDO is carried out with the long-short-term reference scheme, the average BD-BR reductions of Y component are 28.20% and 18.34% for the surveillance videos and the conference videos, respectively. It indicates that employing the enhanced frame as the long-term reference can significantly lift the coding performance. That is, the effect of the enhanced frame has been further improved by the long-short-term reference scheme.

The R-D curves of the proposed two schemes, as well as the anchor, are shown in Figure 6. It shows that both of the two schemes can achieve significant improvements on the anchor for a wide range of bit-rates. Furthermore, the PEF-LSTR scheme outperforms the PEF-dHPS significantly. This also shows the efficiency of the long-short-term reference scheme.

E. COMPARED WITH THE STATE-OF-THE-ART METHODS

The proposed PEF-LSTR method is also compared with the recent works [23] and [24]. The results are displayed in Table 4. It shows that, when encoding the surveillance videos, the average BD-BR reductions (on the Y component) of [23] and [24] are 35.91% and 38.80%, respectively. In that case, the average BR-BR reduction of the proposed method is 28.20%, i.e., the two compared methods are more efficient than the proposed method. However, when encoding the conference videos, the average BD-BR reductions (on the Y component) of [23] and [24] are only 1.15% and 5.27%, respectively. In that case, the proposed algorithm can save 18.34% of coding bits on average. i.e., the proposed PEF-LSTR method significantly outperforms the compared methods.

The two compared methods are the background modeling-based schemes. In these methods, in order to reduce the long-term redundancy, the background is modeled as a new frame and encoded with the small QP values. It can be observed that these methods are significantly efficient for coding the surveillance videos. However, the performance of coding the conference videos are not so obvious. The major reason is that these methods mainly focus on

stationary backgrounds. There are larger proportions of backgrounds in the surveillance videos than in the conference videos. Generally, in the conference videos, some of the backgrounds are temporally obscured by the foregrounds in motion, such as the heads, arms, and bodies. In addition, the foreground motion may cause image blur [31]. Therefore, it is not easy to obtain a clean background [24]. However, the proposed PEF-based long-term reference scheme is not affected by the temporary occlusion. Both the stationary backgrounds and the foregrounds with small motions can benefit from the proposed scheme. Therefore, the proposed method significantly outperforms the compared methods, while they are using to compress the conference videos.

Moreover, even though the proposed method cannot exceed the compared method for coding the surveillance videos, its average BD-BR savings (Y-BDBR: 28.20%, U-BDBR: 56.44%, V-BDBR: 54.85%) are also significant. It still makes sense that, only the QP offset and the prediction structure are changed in the proposed method, and the other coding tools are unnecessary to be changed, such that it is easy to implement and upgrade the proposed method on existing products.

VI. CONCLUSION

In this paper, a PEF-based long-short-term reference structure is proposed for coding the conference videos and surveillance videos. Since these videos are generally acquired by stationary cameras, the temporal dependency of the videos is higher than that of the ordinary videos. Two schemes are presented to further exploit the long-term temporal dependency of the videos. Firstly, unlike independently optimize the CUs in the default RDO scheme, the PEF-based RDO scheme is presented to globally optimize the coding performance of frames in the long periods. The long-term temporal dependency is described as the error propagation characteristic via a line model. Based on this model, the enhanced frames are periodically set to be encoded with a smaller QP. Secondly, the factors, such as temporary occlusions and quantization errors, are analyzed as the limitations of applying the PEF-based

RDO on the default prediction structure. In order to reduce these limitations, we proposed the PEF-based long-short-term reference scheme. In this scheme, the enhanced frames are employed as the long-term reference picture, such that the long-term temporal dependency can be further mitigated. The proposed method is validated by coding the surveillance videos, as well as the conference videos. It shows significant improvement of the PEF-LSTR scheme on the default HEVC scheme.

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 the Chinese National Natural Science Foundation.

REFERENCES

- [1] J. Pan, Y. Yin, J. Xiong, W. Luo, G. Gui, and H. Sari, "Deep learning-based unmanned surveillance systems for observing water levels," *IEEE Access*, vol. 6, pp. 73561–73571, 2018.
- [2] W. Li, H. Li, Q. Wu, F. Meng, L. Xu, and K. N. Ngan, "Headnet: An end-to-end adaptive relational network for head detection," *IEEE Trans. Circuits Syst. Video Technol.*, to be published.
- [3] J. Gantz and D. Reinsel. (2012). *The IDC Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. [Online]. Available: <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- [4] 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, Dec. 2012.
- [5] K. Ugur 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.
- [6] W.-J. Han 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.
- [7] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, 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.
- [8] Z. Pan, X. Yi, and L. Chen, "Motion and disparity vectors early determination for texture video in 3D-HEVC," *Multimedia Tools Appl.*, vol. 1, pp. 1–18, Nov. 2018.
- [9] Z. Pan, J. Lei, Y. Zhang, and F. L. Wang, "Adaptive fractional-pixel motion estimation skipped algorithm for efficient HEVC motion estimation," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, pp. 12:1–12:19, Jan. 2018.
- [10] J. Lei, J. Duan, W. Feng, N. Ling, and C. Hou, "Fast mode decision based on grayscale similarity and inter-view correlation for depth map coding in 3D-HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 706–718, Mar. 2018.
- [11] J. Xiong, H. Li, Q. Wu, and F. Meng, "A fast HEVC inter CU selection method based on pyramid motion divergence," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 559–564, Feb. 2014.
- [12] 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 Trans. Multimedia*, vol. 16, no. 8, pp. 2141–2153, Dec. 2014.
- [13] J. Xiong, H. Li, F. Meng, Q. Wu, and K. N. Ngan, "Fast HEVC inter CU decision based on latent SAD estimation," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2147–2159, Dec. 2015.
- [14] J. Xiong, X. Long, R. Shi, M. Wang, J. Yang, and G. Gui, "Background error propagation model based RDO in HEVC for surveillance and conference video coding," *IEEE Access*, vol. 6, pp. 67206–67216, 2018.
- [15] L. Zhao, Y. Tian, and T. Huang, "Background-foreground division based search for motion estimation in surveillance video coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
- [16] X. Li and Y. Abudoulikemu, "Background-foreground information based bit allocation algorithm for surveillance video on high efficiency video coding (HEVC)," in *Proc. Vis. Commun. Image Process. (VCIP)*, Nov. 2016, pp. 1–4.
- [17] L. Zhao et al., "Intelligent analysis oriented surveillance video coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2017, pp. 37–42.
- [18] 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 *Proc. Vis. Commun. Image Process. (VCIP)*, Nov. 2012, pp. 1–6.
- [19] J. Xiao, R. Hu, L. Liao, Y. Chen, Z. Wang, and Z. Xiong, "Knowledge-based coding of objects for multisource surveillance video data," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1691–1706, Sep. 2016.
- [20] C. Ma, D. Liu, X. Peng, and F. Wu, "Surveillance video coding with vehicle library," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 270–274.
- [21] K. Yang, F. Chen, D. Liu, Z. Chen, and W. Li, "Surveillance video coding with dynamic textural background detection," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2736–2740.
- [22] X. Zhang, Y. Tian, L. Liang, T. Huang, and W. Gao, "Macro-block-level selective background difference coding for surveillance video," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2012, pp. 1067–1072.
- [23] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 769–784, Feb. 2014.
- [24] 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 Trans. Image Process.*, vol. 23, no. 10, pp. 4511–4526, Oct. 2014.
- [25] L. Yin, R. Hu, S. Chen, J. Xiao, and J. Hu, "A block-based background model for surveillance video coding," in *Proc. Data Compres. Conf.*, Apr. 2015, p. 476.
- [26] C. Chen, J. Cai, W. Lin, and G. Shi, "Surveillance video coding via low-rank and sparse decomposition," in *Proc. 20th ACM Int. Conf. Multimedia*, New York, NY, USA, 2012, pp. 713–716. [Online]. Available: <http://doi.acm.org/10.1145/2393347.2396294>
- [27] C. Chen, J. Cai, W. Lin, and G. Shi, "Incremental low-rank and sparse decomposition for compressing videos captured by fixed cameras," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 338–348, Jan. 2015.
- [28] 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 *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6.
- [29] G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," *IEEE Intell. Syst.*, Apr. 2001.
- [30] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, "The IEEE 1857 standard: Empowering smart video surveillance systems," *IEEE Intell. Syst.*, vol. 29, no. 5, pp. 30–39, Sep./Oct. 2014.
- [31] S. Xie, X. Zheng, W.-Z. Shao, Y.-D. Zhang, T. Lv, and H. Li, "Non-blind image deblurring method by the total variation deep network," *IEEE Access*, vol. 7, pp. 37536–37544, 2019.



JIAN XIONG (M'18) received the Ph.D. degree in signal and information processing from the University of Electronic Science and Technology of China, Chengdu, China, in 2015. In 2014, he was a Research Assistant with the Image and Video Processing Laboratory, The Chinese University of Hong Kong, Hong Kong. Since 2015, he has been an Assistant Professor with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China. He has authored or coauthored over 30 refereed technical papers in international journals and at conferences. His current research interests include image and video coding, computer vision, and machine learning. He is a member of the IEEE Circuits and Systems Society.



XIANZHONG LONG received the B.S. degree in computer science from Henan Polytechnic University, in 2007, the M.S. degree in computer science from Xihua University, in 2010, and the Ph.D. degree from Shanghai Jiao Tong University, in 2014. He is currently an Assistant Professor with the Nanjing University of Posts and Telecommunications. His research interests include computer vision, machine learning, and image processing, specifically on image classification, object recognition, and clustering.



interests include object segmentation visual quality evaluation, interactive segmentation, and salient object detection.

RAN SHI received the B.S. degree in electronic science and technology from the Changshu Institute of Technology, in 2009, and the M.S. degree in signal and information processing from Shanghai University, in 2012, and the Ph.D. degree from the Department of Electronic Engineering, The Chinese University of Hong Kong, in 2016. He has been an Assistant Professor with the School of Computer Science, Nanjing University of Science and Technology, Nanjing, China. His research



Information Engineering, Shenzhen University (SZU), Guangdong, China. He has authored or coauthored over 30 refereed technical papers in international journals and at conferences. His current research interests include a wide range of topics related to image/video compression, transmission and analysis, computer vision, and machine learning. He is a member of the IEEE Circuits and Systems Society. He was a recipient of the Best Thesis Award of Shanghai (Ministry of Education of Shanghai City) and Fudan University (FDU), China, in 2012.

MIAOHUI WANG (S'13–M'16) received the Ph.D. degree from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2015. From 2014 to 2015, he was with the Innovation Laboratory, InterDigital Inc., San Diego, CA, USA. From 2015 to 2017, he was a Senior Research Engineer with The Creative Life (TCL) Research Institute of Hong Kong, Hong Kong, working on computer vision and machine learning. He is currently an Assistant Professor with the College of



nals, including IEEE TIP, EL, and sensors), and conferences, including ICIP, ICASSP, ACCV, and ICPR, in image processing and computer vision. His research interests include computer vision and pattern recognition. He serves as a TPC Member of many international conferences and a Reviewer for a series of SCI journals, including the IEEE TIP, the IEEE TC, the IEEE TSP, the IEEE CSVT, PR, and *Neurocomputing*.

QUAN ZHOU (M'17) received the B.S. degree in electronics and information engineering from the China University of Geosciences, Wuhan, China, in 1998, and the M.S. and Ph.D. degrees in communication and information system from the Huazhong University of Science and Technology, China, in 2006 and 2013, respectively. He is currently an Associate Professor with the Nanjing University of Posts and Telecommunications. He has published more than 40 articles in top journals,



respectively, with the support from the China Scholarship Council and the Global Center of Education. From 2012 to 2014, he was supported by the Japan Society for the Promotion of Science Fellowship as a Postdoctoral Research Fellow with the Wireless Signal Processing and Network Laboratory (Prof. Adachi's Laboratory). From 2014 to 2015, he was an Assistant Professor with the Department of Electronics and Information System, Akita Prefectural University. Since 2015, he has been a Professor with the Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China. He is currently involved in research of compressive sensing and deep learning for physical layer signal processing. He received four best paper awards in international conferences, ICNC 2018, ICC 2017, ICC 2014, and VTC 2014-Spring. He was also selected as a Jiangsu Special Appointed Professor and for the Jiangsu High-level Innovation and Entrepreneurial Talent, the Nanjing Youth Award, and the 1311 Talent Plan of NJUPT. He was an Editor of *Security and Communication Networks*, from 2012 to 2016. He has been an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and *KSII Transactions on Internet and Information Systems*, since 2017.

GUAN GUI (SM'17) received the Dr. Eng. degree in information and communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2012. From 2009 to 2012, he was with the Wireless Signal Processing and Network Laboratory (Prof. Adachi's Laboratory), Department of Communications Engineering, Graduate School of Engineering, Tohoku University, as a Research Assistant and a Postdoctoral Research Fellow,

...