Quality-Progressive Coding for High Bit-Rate Background Frames on Surveillance Videos

Shaoge Guo*, Yaowei Wang*, Yonghong Tian†, Peiyin Xing† and Wen Gao†
* School of Information and Electronics
Beijing Institute of Technology, Beijing, China, 100081
Email: guoshaoge0706@163.com, yaoweiwang@bit.edu.cn
† National Engineering Laboratory for Video Technology
School of EE & CS
Peking University, Beijing, China, 100871
Email: yhtian@pku.edu.cn, xingpeiyin@gmail.com, wgao@pku.edu.cn

***Abstract*—A remarkable compress performance was achieved in the surveillance video coding when high-quality reconstructed background frames were referenced in long-term. However, the high-quality reconstructed background frames lead to bursting bit-rate peaks in the video transmission, which may cause obvious communication delay or buffer overflow. In order to address this problem, the paper proposes a quality-progressive coding algorithm for smoothing the bursting peaks caused by high-quality reconstructed background frames. Instead of a single high-quality reconstructed background frame, we divide the background frame into a set of frames, which include the basic reconstructed background frame of normal-quality (basic part) and a series of reconstructed residual frames (residual part), for transmission. Moreover, the modeled background frame and the residual frames should be encoded into the bit-stream and transmitted every several frames, and the coding bits of two frame types above should be limited to the target range of coding bits, which is based on the channel capacity, so that avoids the bursting bit-rate peak and the transmission delay. Background frames are reconstructed by summing up the basic part and reconstructed residual frames one by one, and the last reconstructed background frame becomes a high-quality reconstructed background frame. Of course, each reconstructed background frame act as a prediction reference for its subsequent frames. Experimental results on an opening dataset, PKU-SVD-A, show that the proposed approach can smooth the bit-rate of high-quality reconstructed background frame in surveillance video coding, and achieve 0.57% bit-rate saving on average compared with HEVC-S.**

Keywords*—*Surveillance video coding; background modeling; quality-progressive coding; HEVC

I. Introduction

It was reported that about 20,000 surveillance cameras had been installed in China till 2013. With such a massive amount of cameras, an unimaginable huge volume of video data is generated day and night. This presents a grand challenge for saving and transmitting the huge volume of video data efficiently and effectively. In the past 10 years, video coding technologies have made a successive progress and recent studies on surveillance video coding prove that background based video coding method can further improve the video compress efficiency of surveillance videos generated by stationary surveillance cameras [1]. In background frame (BG-frame) based methods, the long-term referenced BG-frame is often a static image without moving objects, which can be modeled, captured or manually segmented from videos. Moreover, the quality of BG-frame is higher, the compressing performance is better. BG-frame based coding usually adopts the framework shown in Fig.1. In such a framework, the background frames are always encoded into the video bit-stream with high quality to act as good prediction references and transmitted to the decoder. Though the average bit-rate of BG-frame based coding method is lower than that of generic coding methods, the high-quality reconstructed background frame will lead to a bursting bit-rate peak, where the real-time bit-rate is abruptly much higher than the average bit-rate. Comparing with the methods in [2], [3] and [4], the coding and modeling methods in [1] [5] achieved a significant compress performance with low memory cost and computing complexity for surveillance videos. But the bursting bit-rate peak of high-quality reconstructed BG-frame still is its bottleneck to apply in the real surveillance scenario.

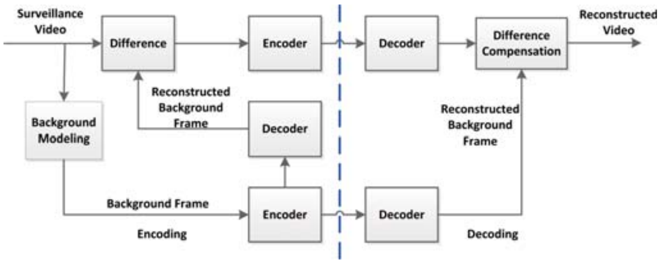

Fig. 1. The framework of background modeling based coding method

As the BG-frame is always an intra-coded frame, we can determine whether the bit-rate is abruptly bursting by comparing the coding bits of the BG-frame with that of I-frames. There is no problem for the first BG-frame since the first BG-frame utilizes the first intra-coded frame (I-frame) as a prediction reference with normal quantization parameters (QP). A smaller quantization parameter means a higher quality for reconstrcuted frames in the video coding. Let $QP_{BG\text{-}frame}$ denote the QP of BG-frame and $QP_{I\text{-}frame}$ be the QP of the I-frame. Practical applications for surveillance video coding proved that it was a good trade-off between compress performance and the bit-rate of the BG-frame when $QP_{BG\text{-}frame}$ was equal to $QP_{I\text{-}frame}-10$. In table I, we compare the coding

TABLE I.
COMPARISON OF CODING BITS OF A BG-FRAME WITH $QP_{I\text{-}frame} - 10$ AND THOSE OF AN I-FRAME WITH $QP_{I\text{-}frame}$

Sequence	$QP_{\text{I-frame}}$	I bits	BG bits	BG/I
Campus(720 × 576)	22	415240	676704	1.63
	27	239016	444944	1.86
	32	136560	282192	2.07
	37	76272	160640	2.11
Indoor(1280 × 720)	22	462744	976352	2.11
	27	266704	600560	2.25
	32	151888	356168	2.34
	37	88664	183836	2.07
Lake(1920 × 1080)	22	2419920	4636056	1.92
	27	1471176	3103672	2.11
	32	847208	1966424	2.32
	37	442664	1107184	2.50
Average	_	_	_	**2.11**

bits of BG-frames at $QP_{I\text{-}frame} - 10$ with those of I-frames at $QP_{I\text{-}frame}$ for three typical sequences of different resolutions. It can be easily observed that the coding bits of BG-frame are nearly 2.11 times of the coding bits of I-frames on average. That means the decoder needs more buffer or will suffer the transmission delay. Moreover, the coding performance will be obviously declined when the BG-frame is engaged in the rate control.

To address this problem, this paper proposes a quality-progressive coding (QPC) method to smooth the abrupt bit-rate peaks caused by the high-quality reconstructed BG-frame. Instead of one single high-quality reconstructed BG-frame, we introduce a set of frames, which includes one reconstructed BG-frame of normal-quality (basic part, the same quality as an reconstructed I-frame) and a set of reconstructed residual frames (residual part), for video stream transmission. Besides, coding bits of the modeled BG-frame and residual frames are limilted to the range according to the channel capacity. The BG-frame is reconstructed by summing up the basic part and frames in the residual part one by one. The first reconstructed BG-frame is of normal-quality like a reconstructed I-frame, and the quality of reconstructed BG-frames increases a bit when a residual part is added. It means that the reconstructed BG-frames are quality-progressive and the reconstructed BG-frame becomes a high-quality reconstructed BG-frame at last. Also, each quality-progressive reconstructed BG-frame acts as a prediction reference for its subsequent frames. Experimental results on an opening dataset, PKU-SVD-A, show that the proposed approach can solve the problem of bursting bit-rate peak, and achieve 0.57% bit-rate saving on average.

The rest of the paper is organized as follows: Section 2 describes the proposed approach in detail. Experimental results are reported and analyzed in Section 3. Finally, conclusions are given in Section 4.

II. THE PROPOSED METHOD

The proposed QPC approach solves the bit-rate bursting problem caused by high-quality reconstructed BG-frames in BG-frames based surveillance video coding. The features of this approach are three-fold: first, a set of frames are introduced to decrease the real-time bit-rate in video transmission; second, the reconstructed BG-frames are quality-progressive; finally, each updated reconstructed BG-frame is a prediction reference for its subsequent frames.

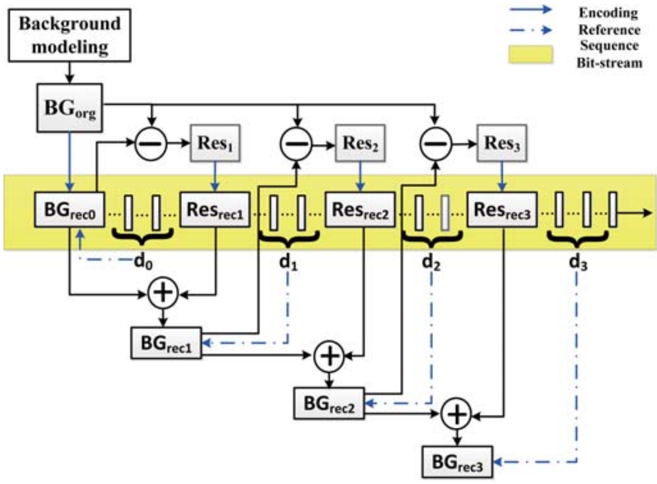

Fig. 2. A four-layer QPC framework

A. The Framework of Encoding

A four-layer QPC framework for the BG-frame is shown in Fig.2. Obviously, the framework can easily scale to N layers. The original BG-frame (BG_{org}) is modeled with some selected original video frames that depends on the modeling approach and then encoded into the video stream with $QP_{I\text{-}frame}$ to avoid the bit-rate bursting with coding bits limited to the target range. This BG-frame (BG_{rec0}) is the basic part in our hierarchical frames. BG_{rec0} is a critical prediction reference frame for its subsequent frames. In the four-layer QPC framework, the modeled BG-frame is divided into BG_{rec0} and $Res_{rec(i)}$ (i=1,2,3), where $Res_{rec(i)}$ is a reconstructed residual frame. The difference frame between BG_{org} and $BG_{rec(i-1)}$ is the residual frame ($Res_{(i)}$), which is encoded for transmission within the coding bit limit. The sum of $BG_{rec(i-1)}$ and $Res_{rec(i)}$ is $BG_{rec(i)}$, a quality-progressive reconstructed BG-frame, which is a prediction reference frame for the following d_i frames. With such a framework, the quality of reconstructed BG-frame becomes higher and higher. Finally, the quality of the last reconstructed BG-frame is better than that of BG_{rec} in [1]. Therefore, the problem of bit-rate bursting can be solved and the compression performance can be further improved with our approach.

B. Background Modeling and Updating

The proposed method utilizes the low complexity Segment-and-Weight based Running Average algorithm [5] and S-GOP [1] as the background modeling and updating method. Obviously, the proposed method is independent of the modeling and updating methods. Theoretically, the proposed approach can be applied in all kinds of BG-frame based surveillance coding methods to overcome the bit-rate bursting caused by the high-quality reconstructed BG-frame.

C. The Residual Frame Encoding and Updating

In the framework, the basic reconstructed BG-frame (BG_{rec0}) acts as a prediction reference frame for the subsequent d_0 frames. Therefore the i^{th} residual frame can be obtained as following:

$$Res_{(i)} = BG_{org} - BG_{rec(i-1)}, (i \geq 1) \quad (1)$$

Then the residual frame is encoded into the bit-stream and its coding bits can be adjusted by changing QP to be within the target range. In this paper, the target range of coding bits is based on the coding bits of intra-coded frames. Hence, the updated reconstructed BG-frame can be computed by (2), which is not only a reference frame for the subsequent d_i frames, but the input of (1) to generate the next residual frame.

$$BG_{rec(k)} = BG_{rec0} + \sum_{i=1}^{k} \mathrm{Res}_{rec(i)} \quad (k \geq 1, d_3 \gg d_0, d_1, d_2) \tag{2}$$

It should be noted that the pixel values of a residual frame may be negative. To make its easier implementation on encoder, we add them an offset to assure the pixel values are positive. For such a goal, the offset's greatest lower bound should be 255. This usually needs 9-bit coding for each pixel. If we can keep the pixel values with maximum absolute values from residual frames positive by using an offset, other pixel values will also be positive. Whereas, the distribution of pixel values is not even. As shown in Fig.4(a), maximum absolute values of Y are increasing but not more than 80 when QP is increased. U and V has the same phenomenon. Therefore, we select an empirical offset value, 128. Those pixels, whose values are negative after adding the offset, are treated as outliers and set to 0. Similarly, pixels whose values are greater than 255 are set to 255. By doing this, we can encode the residual frames with 8-bit depth, and assure all pixel values are positive.

Practically, the pixel values of $Res_{(i)}$ should be calculated by (3) and then transferred,

$$RES_{(i)}(x,y) = Clip(0, 255, \mathrm{Res}_{(i)}(x,y) + \mathit{offset}) \tag{3}$$

, where $Res_{(i)}(x,y)$ is the original pixel value at position (x,y) in the residual frame $Res_{(i)}$, $RES_{(i)}(x,y)$ is the pixel value after the offset at position (x,y) in the residual frame $Res_{(i)}$. Therefore, at the decoder, the offset should be subtracted when a residual frame is received and reconstructed. The pixel value of the reconstructed residual frame is computed by (4),

$$Res_{rec(i)}(x,y) = RES_{REC(i)}(x,y) - \mathit{offset} \tag{4}$$

, where $Res_{rec(i)}(x,y)$ is the pixel value without offset. The quality-progressive reconstructed BG-frames can be computed by (5),

$$BG_{rec(i)}(x,y) = Clip(0, 255, Res_{rec(i)}(x,y) + BG_{rec(i-1)}(x,y)) \tag{5}$$

, where $BG_{rec(i)}(x,y)$ is the pixel value at position (x,y) of the reconstructed BG-frame. $BG_{rec(i)}$ is used as a prediction reference for the following d_i frames and the input for generating the next residual frame.

Obviously, we can continuously update the reconstructed BG-frame until its quality is high enough (i.e., the quality of reconstructed BG-frame at $QP_{I\text{-}frame} - 10$). The quality could be evaluated by comparing values of PSNR between each reconstructed BG-frame and the modeled BG-frame.

TABLE II.
THE MODIFICATION OF SYNTAX ELEMENT ABOUT PIC_TYPE

Pic_type	Slice_type values that maybe present in the coded picture
0	I
1	P, I, **U**, **BG**
2	B, P, I, **U**, **BG**

TABLE III.
THE MODIFICATION OF SYNTAX ELEMENT ABOUT SLICE_TYPE

Slice_type	Name of slice_type
0	B(B slice)
1	P(P slice)
2	I(I slice)
3	**BG(BG slice)**
4	**U(U slice)**

D. The Modification of Syntax

At the decoder, this method needs to define two new frame types, the reconstructed BG-frame (BG) and the reconstructed residual frame (U). In the standard specification [6], the table of interpretation of pic_type, U and BG should be added if pic_type is equal to 1 or 2. Table II and III shows the modification of syntax about pic_type and slice_type, respectively. If the slice_type is 3, the decoding frame will be a basic reconstructed BG-frame, and if the slice_type is 4, the decoding frame will be a reconstructed residual frame.

III. EXPERIMENTAL RESULTS

A. Experiment Settings

Apparently, there is no background modeling or long-term reference frames in HM12.0 while HEVC-S aiming at surveillance videos adopts the long-term reference mechanism [1] based on the background modeling method of SWRA in [5]. The software HEVC-S achieves a significant compression efficiency but hasn't dealt with the bursting bit-rate peaks. On the improved software with QPC, the experiment adopts surveillance videos from the dataset of PKU-SVD-A [7] [8] which are shot by stationary cameras. For the real-time transmission of surveillance videos, the experiment is only done under the low delay configuration [9]. We use six test sequences of 1020 frames, and they are shown in Fig.3. In this experiment, d_0, d_1 and d_2 are set as 10 and at the same time, the condition in (6) must be guaranteed,

$$d_0 + d_1 + d_2 + d_3 = GOP - Length \tag{6}$$

, where GOP-Length is the update period of BG-frame in [1] (GOP-Length = 900 in our experiment). We employ $QP_{I\text{-}frame} - 10$ as QP_{BG} in HEVC-S as the anchor. The number of residual frames is 3, which makes the ultimate reconstructed BG-frame of higher quality than that with $QP_{I\text{-}frame} - 10$ in HEVC-S, and $offset$ is set as 128 in this experiment. In addition, BG_{org} is encoded into the bit-stream with $QP_{I\text{-}frame}$. QP values for 3 layers of residual frames are respectively $QP_{I\text{-}frame} - 5$, $QP_{I\text{-}frame} - 9$ and $QP_{I\text{-}frame} - 11$, which keep coding bits of the residual frames within the range. The upper limit of coding bits of the BG-frame and each residual frame we set are 1.1 times of coding bits of an I-frame, which prevents the bursting bit-rate peaks

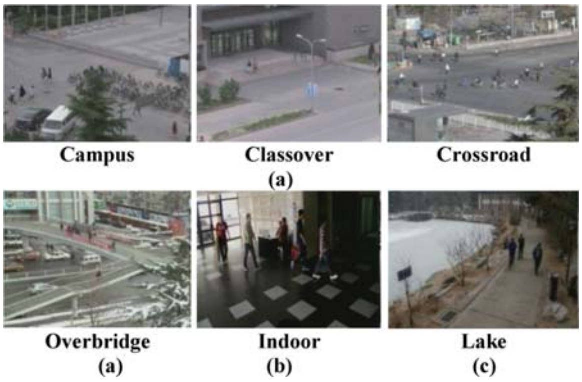

Fig. 3. (a) four 576p (720 × 576) sequences; (b) one 720p (1280 × 720) sequence (c) one 1080p (1920 × 1080) sequence

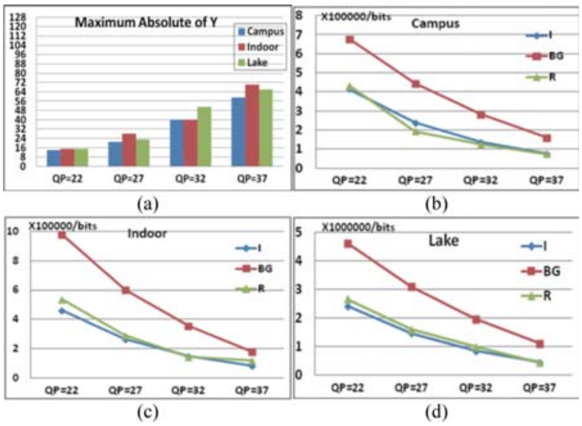

Fig. 4. (a) The maximum absolutes of Y component in residuals from sequences of Campus (blue), Indoor (red) and Lake (green); (b),(c),(d) Comparison of coding bits of I-frame with $QP_{I\text{-}frame}$, BG-frame with $QP_{I\text{-}frame} - 10$ and maximum coding bits of Residual frame with $QP_{I\text{-}frame} - 5$, $QP_{I\text{-}frame} - 9$ or $QP_{I\text{-}frame} - 11$

and the transmission delay. The other test conditions are the same as those of HEVC.

B. Experimental Results

The experimental results are shown in Fig.4 and Table IV. The proposed method gets the final reconstructed BG-frame of better quality than that of HEVC-S, which can be evaluated by comparing the values of PSNR between each reconstructed BG-frame and the original modeled BG-frame. Compared with HEVC-S, the proposed method solves the problem of bursting bit-rate peaks and achieves compression performance shown in Table IV. Besides, this experiment also gives a presentation about coding bits of I-frames and the maximum coding bits of residual frames from 3 representative sequences in Fig.4 (b),(c),(d) which shows coding bits of residual frames are in the range and demonstrates the bursting bit-rate peaks is smoothed.

IV. Conclusion

The proposed method puts forwards a quality-progressive coding method for BG-frames aiming at the bursting bit-rate peaks of background frames. Firstly, the modeled BG-frame is encoded with $QP_{I\text{-}frame}$ and we can achieve the

TABLE IV.
THE COMPARISON OF PERFORMANCE

Resolution	**Sequence**	**QPC VS HEVC-S**	
		BD Rate	BD PSNR
720×576	campus	-0.81%	0.023dB
	classover	-0.39%	0.013dB
	crossroad	-0.48%	0.017dB
	overbridge	-0.93%	0.023dB
1280×720	Indoor	-0.34%	0.026dB
1920×1080	Lake	-0.49%	0.025dB
Average		**-0.57%**	**0.021dB**

basic reconstructed BG-frame. Then a residual frame could be calculated by employing the modeled BG-frame and the basic reconstructed BG-frame, which should be encoded into the bit-stream with coding bits limited to 1.1 times of coding bits of I-frame and transported. Meanwhile, the reconstructed BG-frame serves as a prediction reference for the subsequent frames. The other reconstructed BG-frames could be updated every several frames as stated above. This method keeps the original performance as well as solves the problem of the bursting bit-rate peaks. Compared with the coding method adopting small quantization parameters to directly get a high-quality recosnstructed BG-frame, the proposed method saves 0.57% bit-rate and avoids the burst bit-rate peak. For the future work, we will focus on the residual coding method under the condition of the rate control for surveillance videos.

Acknowledgment

This work is partially supported by grants from the National Basic Research Program of China under grant 2015CB351806, and the National Natural Science Foundation of China under contract No. 61171139, No. 61390515 and No. 61471042.

References

[1] X. Zhang, L. Liang, Q. Huang, Y. Liu, T. Huang, and W. Gao, "An efficient coding scheme for surveillance videos captured by stationary cameras," in *Visual Communications and Image Processing 2010*. International Society for Optics and Photonics, 2010, pp. 77 442A–77 442A.

[2] M. Tiwari and P. C. Cosman, "Selection of long-term reference frames in dual-frame video coding using simulated annealing," *Signal Processing Letters, IEEE*, vol. 15, pp. 249–252, 2008.

[3] A. Hakeem, K. Shafique, and M. Shah, "An object-based video coding framework for video sequences obtained from static cameras," in *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005, pp. 608–617.

[4] M. Paul, W. Lin, C. T. Lau, and B. Lee, "Video coding using the most common frame in scene," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 734–737.

[5] X. Zhang, Y. Tian, T. Huang, and W. Gao, "Low-complexity and high-efficiency background modeling for surveillance video coding." in *VCIP*, 2012, pp. 1–6.

[6] R. H. ITU-T, "265 (04/13)," *Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services–Coding of Moving Video, High Efficiency Video Coding, Online: http://www. itu. int/rec/T-REC-H*, pp. 265–201 304.

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

[8] W. Gao, Y. Tian, T. Huang, S. Ma, and X. Zhang, "IEEE 1857 standard empowering smart video surveillance systems," 2013.

[9] F. Bossen and H. Common, "Test conditions and software reference configurations, jct-vc doc," *L1100, Jan*, 2013.