Rate Model Considering Nontexture Bits for High Efficiency Video Coding

Zehu Huang[1,2], Huizhu Jia[2*], Xiaodong Xie[2]
[1]SECE of Shenzhen Graduate School, Peking University, 518055, Shenzhen, China
[2]National Engineering Laboratory for Video Technology, Peking University, Beijing, China
zehu_huang@pku.edu.cn

***Abstract*—One of the challenges in video rate control lies in accurately achieving the target bitrate by determining coding parameter set. In HEVC, nontexture bits become more significantly important, due to the adoption of more coding tools compared to H.264 standards. In this paper, we propose an accurate rate model to estimate the size of nontexture bits accurately. In addition, parameter update conditions and a model failure detection scheme are introduced to avoid large deviations of the frame size from the target size. In comparison to the original rate model without nontexture bits estimation, the proposed method can achieve better video quality with a PSNR gain up to 0.47 dB and reduce the mismatch between actual frame bits and target ones by up to 64.3%.**

Keywords—High Efficiency Video Coding; rate control; linear rate-quantization model; nontexture rate model.

I. Introduction

HEVC (High Efficiency Video Coding) [1] is the latest international video coding standard and aims at achieving a bit rate reduction of up to 50% compared with H.264/AVC. It is developed by JCT-VC (Joint Collaborative Team on Video Coding), based on conventional block-based hybrid video coding framework.

In addition to coding efficiency, rate control is also an important issue in video encoder systems, especially for real time communication applications. In many video services, rate control plays an important role in delivering a compressed video stream under a certain channel bandwidth restriction. Without rate control, buffer overflow and underflow may occur, which will cause frame skipping and wastage of channel bandwidth, respectively. Therefore, rate control has become a hot research area in the field of video communication.

In HEVC, two rate control proposals have been adopted, which are JCTVC-H0213 [2] and JCTVC-K0103 [3]. JCTVC-H0213 proposes a pixel based unified rate-quantization (R-Q) model but has significant performance loss, due to its inaccurate rate model and the high fluctuation of nontexture bits. Therefore, JCTVC-K0103 using an R-lambda model is adopted and implemented in the HEVC reference software.

In our experiment, we find that the R-lambda model may sometimes fail to achieve the target frame size accurately when applied to complex video sequences, such as movies and animations. Because the R-lambda model don't do any preanalysis of the current frame, and its parameter is updated based on characteristics of previously encoded frames, it might suffer from relatively large errors in terms of rate control accuracy when scene cut or large motion occurs.

In [4], Merritt and Vanam introduce the rate control algorithm in x264 [5], which has been used in many popular applications like ffmpeg [6]. The rate control in x264 employs a linear R-Q model and the SATD (Sum of Absolute Hadamard Transformed Differences) as a measure of frame complexity. Compared with other rate models, it can provide more accurate bitrate estimation and adapt more quickly to the characteristics of the video clips containing complex texture and large motions. Due to its high performance, it has been migrated to x265 [7], which is the most popular open source implementation of HEVC. However, when it is applied to HEVC, improvement need to be made because of the large amount of nontexture bits. In HEVC, nontexture bits, which includes quadtree split information for CU and TU, motion vectors, prediction information etc., become significantly important due to the adoption of various coding tools compared to previous video coding standards. In addition, as shown in Fig. 1, the amount of these nontexture bits will fluctuate from frame to frame, and it cannot be simply estimated. In order to improve the rate model accuracy in HEVC, accurate estimation of nontexture bits is very important.

In this paper, we propose an effective method to estimate the amount of nontexture bits. In addition, parameter update conditions and a model failure detection scheme are introduced to avoid large deviations of the frame size from the target size. Experimental results show that the proposed method can achieve better video quality and improved bitrate accuracy compared to the original rate model in x265.

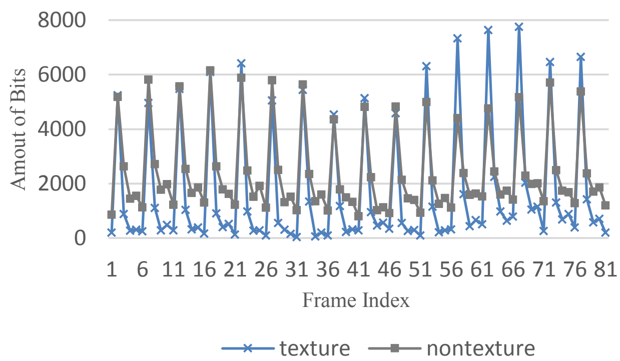


Fig. 1 Fluctuation of the amount of texture bits and nontexture bits for the FourPeople test sequence (1280x720, encoded with QP = 40).

*the corresponding author, Huizhu Jia is with Peking University, also with Cooperative Medianet Innovation Center and Beida (Binhai) Information Research.

The rest of this paper is organized as follows. In section II, nontexture and texture rate models are proposed with respect to QP and frame complexity. Section III presents parameter update conditions and a model failure detection method used in our rate models. Experimental results are presented in Section IV. Section V concludes the paper.

II. Nontexture And Texture Rate Models For HEVC

There have already been a lot of researches on rate models and their applications to rate control. Most of the existing schemes focus mostly on the rate used for coding texture information, such as TM5 for MPEG-2 [8], TMN8 for H.263 [9], and VM8 for MPEG-4 [10]. In these standards, the amount of nontexture bits is much less than texture bits and can be neglected. In HEVC, the proportion of nontexture bits is significantly higher compared to previous standards because of the adoption of many new coding tools. At the same time, the energy of residual signal is reduced by more efficient prediction schemes. As a result, the nontexture bits occupy a higher percentage of the total bits and become even larger than texture bits, especially at low bit rate. Considering the increased proportion and high fluctuation of nontexture bits, we set up rate models for nontexture bits and texture bits respectively. Specifically, the amount of total frame bits is modeled as

$$R = R_{nonTex} + R_{tex} \quad (1)$$

where R_{nonTex} is the estimation of header bits including the split information of quadtree CU and TU, the coded block flags, the motion vectors etc. R_{tex} is the estimation of texture bits.

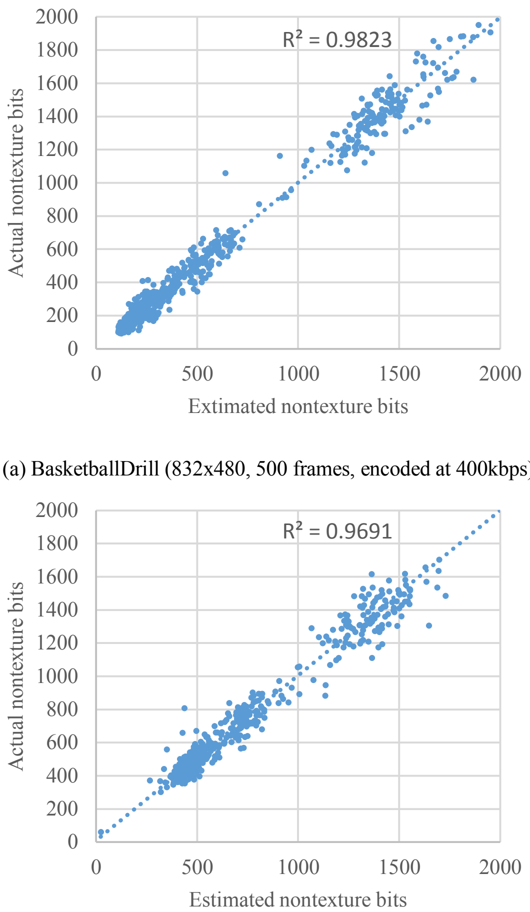
A. Rate Modeling for Nontexture Bits

Most of the conventional rate control algorithms estimate nontexture bits by simply averaging the nontexture data from a few previously encoded frames [11], [12], [13]. However, it cannot handle the fluctuation of nontexture bits over frames. In [14], nontexture bits are predicted by using the number of nonzero differential motion vector elements, which may not be accurate for HEVC with complicated coding structure and various coding tools.

Since the QP used in quantization affects R-D optimization, including the depth of quadtree split for CU and TU, PU partition, intra prediction mode and motion information, it is not surprising that nontexture bits are dependent on QP. In [15], He and Wu find that not only the amount of texture bits but also the amount of nontexture bits is a linear function of $(1-\rho)$, which is defined as the percentage of nonzero transformed coefficients after quantization. However, it is difficult to map ρ to QP due to the complicated quantization scheme. The proportion of nonzero quantized transformed coefficients implies that the higher the frame complexity is, the more bits will be produced. In other words, the amount of nontexture bits is proportional to the frame complexity. Motivated from these observations, we model the nontexture bits as

$$R_{nonTex} = \theta \cdot SATD/QP \quad (2)$$

where R_{nonTex} is an estimation of nontexture bits, SATD denotes current frame complexity, and θ is a model parameter. The initial value of θ is set to 0.1 and will keep updating with the encoding proceeding.



(a) BasketballDrill (832x480, 500 frames, encoded at 400kbps)

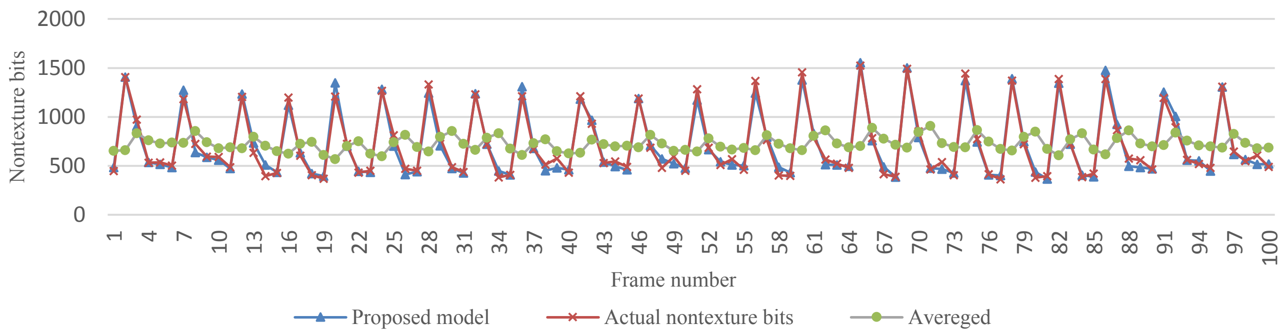
(b) BQMall (832x480, 600 frames, encoded at 400kbps)

Fig. 2 Correspondence of actual nontexture bits and estimated nontexture bits.

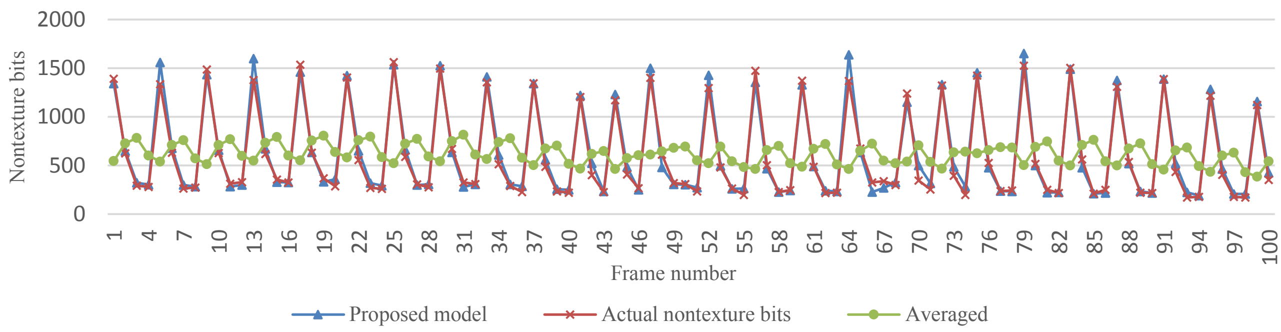
To verify the accuracy of the proposed nontexture rate model, we have tested the rate model on various test sequences. Fig. 2 shows the relation between actual nontexture bits and estimated nontexture bits for BasketballDrill and BQMall sequences. We use correlation coefficient which is an indicator from 0 to 1 to measure how closely the approximated linear function is to the actual data. It is shown that the correlation coefficients are above 0.96. In other words, our nontexture rate model is quite effective. Fig. 3 shows performance comparison of nontexture bit estimations for the proposed nontexture rate model and an average nontexture rate model. It is clearly that the proposed nontexture rate model outperforms the average model a lot in estimation accuracy for nontexture bits. Thus, the linear nontexture rate model is used in this work.

B. Rate Modeling for Texture Bits

For HEVC rate modeling, [16] takes well into account the different statistical characteristic of transformed coefficient residuals by using multiple Laplacian PDFs for different coding unit (CU) categories in various depth levels of HEVC. However, statistical characteristics of residuals using the weighted moving average from the previous frames, are not accurate when residuals change abruptly due to large motions or scene changes. Instead of estimating characteristics from previous frames, x265 runs a fast motion estimation algorithm over a half-resolution version of each frame and uses the SATD of residuals as the



(a) BasketballDrill (832x480, encoded at 400kbps)



(b) BQMall (832x480, encoded at 400kbps)

Fig. 3 Estimation of nontexture bits.

characteristic. In our work, texture rate model is based on x265 and can be described as

$$\mathrm{R}_{tex} = \alpha \times (SATD + \beta)/qscale \quad (3)$$

where R_{tex} is an estimation of texture bits, α and β are two model parameters, set to 1, 0 respectively. The relation between qscale and QP is

$$\text{qscale} = 0.85 \times 2^{(QP-12)/6} \quad (4)$$

It must be pointed out that (3) is used to estimate total frame bits in x265, which may sometimes suffer from large mismatch between actual frame bits and target ones due to the high fluctuation of texture and nontexture bits. In our experiments, we observe that after encoding a series of relatively static or smooth frames, in which most of the quantized coefficients are near zero and the proportion of nontexture bits is much larger than that of texture bits, the model parameter α will be updated based on the characteristic of nontexture bits rather than that of total frame bits. In this situation, α will be very small and no longer valid for future frames which contain large texture bits. To address the above mentioned issue, we setup rate models for texture and nontexture bits separately to achieve more precise rate estimation.

III. Parameter Update Condition And Model Failure Detection Method

It is observed from our experiments that the linear rate models for texture and nontexture bits may lose its accuracy at extremely low bitrates. However, at a normal bitrate range, they are fairly accurate. To address this problem, we update the parameters of texture and nontexture rate models if the following conditions are met:

$$R_{actualTex} > c_1 \times R_T/f \quad (5)$$

$$R_{actualNontex} > c_2 \times R_T/f \quad (6)$$

where $R_{actualTex}$ is actual texture bits, $R_{actualNontex}$ is actual nontexture bits, both are generated by the entropy module, R_T is the target bitrate, f is the frame rate, c_1 and c_2 are constants, set to 0.075 and 0.025 respectively.

Besides, we find that the linear rate models for texture and nontexture bits may fail for sequences containing abrupt scene changes or large motions, resulting in severe bitrate mismatch and quality degradation. To reduce the deviation of actual frame bits from target ones, we use (7) to detect such scene changes or large motions.

$$\text{SATD}[i] > \gamma \times blurredCplx[i-1] \quad (7)$$

$$blurredCplx[i-1] = sum[i-1]/count[i-1] \quad (8)$$

$$\text{sum}[i] = 0.5 \times sum[i-1] + SATD[i] \quad (9)$$

$$\text{count}[\mathrm{i}] = 0.5 \times count[i-1] + 1 \quad (10)$$

where SATD$[i]$ denotes frame complexity of current frame i, γ is a fixed factor and is empirically set to 4, $blurredCplx[i-1]$ represents blurred complexity of previously encoded frames. When scene change or large motion is detected, parameters of the texture and nontexture rate model will be reset to initial values.

IV. Experimental Results

To evaluate the performance of the proposed rate model (1), we implemented it in x265's rate control scheme. The encoder is configured to conform to the main profile. The motion search range is 57 and the GOP (Group of Pictures) size is set to 120. For each GOP, the first frame is encoded as I frame and the remaining frames are encoded as P frames. We randomly select some hot videos as a test set, which contain fast moving objects and abrupt scene changes.

We compare the proposed rate model (1) with the original R-Q model (3) in x265, which is used to estimate total frame bits without considering nontexture bits and model failure detection. For fair comparison, the proposed rate model and the original rate model both adopt the frame level rate control.

A. Video Quality in PSNR

Table I shows a subset of our test results in terms of PSNR. As shown in the table, the proposed rate model can achieve better PSNR than the original model and the largest PSNR improvement is up to 0.47 dB. Fig. 4 shows RD performances comparison between the proposed rate model and the original rate model for some clips. It can be seen that the proposed rate model has better RD performance than the original rate model. This is because the proposed rate model, considering accurate estimation of nontexture bits and model failure detection scheme, can control the frame size more accurately.

B. Bit Accuracy of Rate Control

From Table I, one can also see that the actual bitrate produced by the proposed model is closer to the target bitrate in most cases when compared to the original rate model.

To further compared the rate model accuracy, Table II presents the average deviation of the frame size from the target bit budget for each clip, which is calculated with (11).

$$\text{Dev} = \left(\sum_i \left(|R_A[i] - R_T[i]|\right)/R_T[i]\right)/N \qquad (11)$$

where $R_A[i]$ and $R_T[i]$ are the frame size and target bit budget of frame i respectively, and N is the total number of encoded frames in a video.

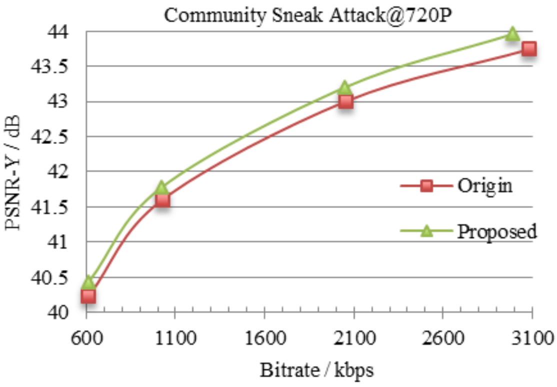
From Table II, it can be observed that the proposed rate model has the smaller deviation than the original rate model and the largest reduced deviation is up to 64.3%. In other words, the proposed rate model can achieve the target bit budget more accurately and make full use of the channel bandwidth.

V. Conclusion

In this paper, we proposed an accurate rate model to estimate the size of nontexture bits accurately. In addition, parameter update conditions and a model failure detection method are introduced to avoid large deviations of the frame size from the target size. Experimental results show that the proposed method can achieve better video quality with a PSNR gain up to 0.47 dB and reduce the mismatch between actual frame bits and target ones by up to 64.3%, when compared with the original rate model without nontexture bits estimation.

TABLE I. Experimental Results on Rate Control

Sequences	Target BR (kbps)	Original		Proposed		
		BR (kbps)	PSNR (dB)	BR (kbps)	PSNR (dB)	PSNR Gain (dB)
Master Chef @ 720P	600	609.89	43.03	607.54	43.27	0.24
	1000	1016.70	44.63	1013.17	44.86	0.23
	2000	2014.40	46.36	2015.55	46.53	0.18
	3000	2991.36	47.35	2990.97	47.36	0.01
Community Sneak Attack @ 720P	600	615.23	40.24	612.15	40.43	0.19
	1000	1025.28	41.60	1022.35	41.78	0.18
	2000	2053.56	43.00	2048.47	43.20	0.20
	3000	3078.81	43.74	2990.97	43.96	0.21
The Office Bursting Balloon @ 720P	600	612.64	40.17	598.33	40.32	0.14
	1000	1021.54	41.47	1020.16	41.67	0.20
	2000	2042.81	42.57	2041.79	42.85	0.27
	3000	3064.20	43.16	3063.61	43.36	0.20
The Simpsons @ 576P	600	603.00	37.26	602.63	37.72	0.47
	800	802.94	39.45	801.90	39.85	0.40
	1000	1000.04	41.19	998.90	41.57	0.38
	1200	1189.42	42.56	1188.78	42.89	0.33
Bobs Burgers @ 576P	600	599.11	43.93	599.58	44.40	0.47
	800	774.81	46.37	772.19	46.66	0.28
	1000	923.20	47.94	919.93	48.10	0.16
	1200	1046.96	48.98	1044.89	49.09	0.10



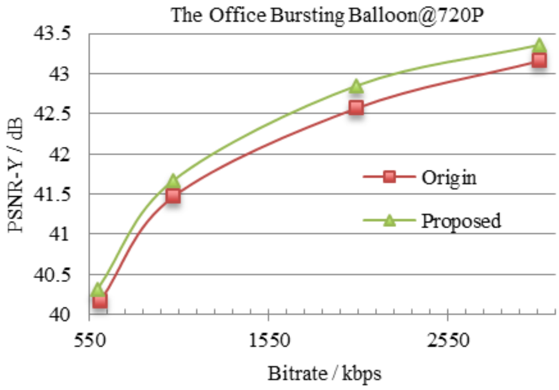


Fig. 4 Selected RD performances

TABLE II. AVERAGE DEVIATION OF THE FRAME SIZE FROM THE TARGET BIT BUDGET

Sequences	Original Dev[%]	Proposed Dev[%]	Reduced Dev[%]
Master Chef @ 720P	23.2	18.5	4.7
Community Sneak Attack @ 720P	66.9	19	47.9
The Office Bursting Balloon @720P	84.6	20.3	64.3
The Simpsons @ 576P	81.5	50.6	30.9
Bobs Burgers @ 576P	80.8	57.7	23.1

ACKNOWLEDGMENT

This work is partially supported by grants from National Key Technology Research and Development Program of the Ministry of Science and Technology of China under contract No. 2014BAK10B00, National High Technology Research and Development Program of China (863 Program) under contract No.2015AA015903, the Major National Scientific Instrument and Equipment Development Project of China under contract No. 2013YQ030967.

REFERENCES

[1] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[2] H. Choi, J. Nam, J. Yoo, D. Sim and I.V. Bajić, "Rate control based on unified RQ model for HEVC," ITU-T SG16 Contribution, JCTVC-H0213, San José, Feb. 2012.

[3] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," ITU-T SG16 Contribution, JCTVC-K0103, Shanghai, Oct.2012.

[4] L. Merritt and R. Vanam, "Improved Rate Control and Motion Estimation for H.264 Encoder," *2007 IEEE International Conference on Image Processing*, San Antonio, TX, 2007, pp. 309-312.

[5] "x264," available: http://www.videolan.org/developers/x264.html

[6] "ffmpeg," available: https://ffmpeg.org/

[7] "x265," available: http://x265.org/

[8] M. ISO/IEC JTC1/SC29/WG11, MPEG-2 Video Test Model 5, Apr.1993.

[9] J. Ribas-Corbera and Shawmin Lei, "Rate control in DCT video coding for low-delay communications," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172-185, Feb 1999.

[10] A. Vetro, Huifang Sun and Yao Wang, "MPEG-4 rate control for multiple video objects," in IEEE Transactions on *Circuits and* Systems for Video Technology, vol. 9, no. 1, pp. 186-199, Feb 1999.

[11] Minqiang Jiang and Nam Ling, "On Lagrange multiplier and quantizer adjustment for H.264 frame-layer video rate control," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 5, pp. 663-669, May 2006.

[12] M. Gao, B. Cizmeci, M. Eiler, E. Steinbach, D. Zhao and W. Gao, "Macroblock level rate control for low delay H.264/AVC based video communication," Picture Coding Symposium (PCS), 2015, Cairns, QLD, 2015, pp. 210-215

[13] J. Y. Chen, C. W. Chiu, G. L. Li and M. J. Chen, "Burst-Aware Dynamic Rate Control for H.264/AVC Video Streaming," in IEEE Transactions on Broadcasting, vol. 57, no. 1, pp. 89-93, March 2011

[14] F. Zhang and E. Steinbach, "Improved ρ-domain rate control with accurate header size estimation," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, 2011, pp. 813-816.

[15] Z. He and D. O. Wu, "Linear Rate Control and Optimum Statistical Multiplexing for H.264 Video Broadcast," in IEEE Transactions on Multimedia, vol. 10, no. 7, pp. 1237-1249, Nov. 2008.

[16] B. Lee, M. Kim and T. Q. Nguyen, "A Frame-Level Rate Control Scheme Based on Texture and Nontexture Rate Models for High Efficiency Video Coding," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 3, pp. 465-479, March 2014