More Probability Estimators for CABAC in Versatile Video Coding

2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP) | 978-1-7281-6896-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICSIP49896.2020.9339374

Sio-Kei Im*
Macao Polytechnic Institute
Macao, China
e-mail: marcusim@ipm.edu.mo

Ka-Hou Chan
School of Applied Sciences
Macao Polytechnic Institute
Macao, China
e-mail: chankahou@ipm.edu.mo

***Abstract*—In the next-generation Versatile Video Coding (VVC), its Context-based Adaptive Binary Arithmetic Coding will have evolved to use linear quantization and multi-probability estimators for prediction, instead of using the finite state table and lookup table in the current standard. This paper highlights the key techniques that involve more probability estimators to enable VVC to potentially achieve a linear quantization representation of the probability state, and a multiplication-free operation can be achieved through a series of bit operations. This article focuses on key techniques involving more probability estimators, so that VVC can linearly achieve a quantized representation of probability prediction and describes the scalability potential for higher accuracy. The BD-RATE gain of the proposed approach is up to 4.0% for all-intra mode and 2.0% for inter mode, the improvement showing that gain is provided in the encoding, and the proposed method can be used in the next generation VVC standard.**

Keywords-versatile video coding; CABAC; probability estimator; VVC test model

I. INTRODUCTION

With the development of video entertainment, and realistic visual experiences in consumer electronic devices, standardization committees are dedicating specific compression tools to this type of content. The High-Efficiency Video Coding standard (HEVC) [1] is a widely used video coding system that can provide good performance and high quality support. Further, the development of the new generation video codec standard, Versatile Video Coding (VVC) [2], is supposedly a universal replacement for HEVC and all of its extensions. The flexibility of the VVC will improve the coding efficiency when compared with HEVC, achieving better frame quality at the same coding bit-rate, and supports enhanced compression efficiency and higher ultra-resolution frames. As more prediction methods are provided, VVC continues to use Rate-Distortion Optimization (RDO) as the best way to determine the optimal coding with lower computational cost [3]. The role of the video encoder is to find the best possible encoding mode to maximize the video quality and minimize the bit rate. To meet those requirements, an entropy coding system, called Context-based Adaptive Binary Arithmetic Coding (CABAC) [4], is widely used for video compression standards like early AVC and current HEVC. In fact, CABAC is a variant of arithmetic coding. It proposes to perform integer bit operations on all floating-point type calculations, and pre-defines the probability estimate into 64 finite states. This concept leads to a good trade-off between complexity and compression efficiency. Especially for the hardware implementation, the finite states can be achieved by lookup table, which can keep the entropy coding multiplication free.

A. Related Work

With the coming of next generation VVC, CABAC has also evolved this. Reviewing AVC and HEVC, there are 64 values (finite states) used to represent an accurate estimation. Each value has assigned one of 64 representative values dividing the range of $[0.01875,0.5]$, and the update of probabilities in the context model is based on this rule:

$$P(t+1) = \begin{cases} \alpha \cdot P(t) & MPS \\ 1-\alpha \cdot \big(1-P(t)\big) & LPS \end{cases} \quad (1)$$

Here MPS and LPS mean the Most Probable Symbol and Least Probable Symbol, respectively, and α is the adaptation rate, which is a constant value equal to $\sqrt[63]{0.01875/0.5} \approx 0.949$ (More detail is given in our earlier paper [5]). In the next generation, VVC introduces a new concept that this adaptation rate becomes dynamic during the process of arithmetic coding [6] [7]. The context model will update according to the value of the currently encoded symbol. The adaptation rate is controlled in parallel by two context models, with its parameters r_0 and r_1 as follows:

$$P_i(t+1) = \begin{cases} \left(1-\frac{1}{2^{r_i}}\right) \cdot P_i(t) & MPS \\ 1-\left(1-\frac{1}{2^{r_i}}\right) \cdot \big(1-P_i(t)\big) & LPS \end{cases} \quad (2)$$

where $i = 0,1$, and r_i are exponentially weighted related to different coding modes. The updated probability will be the average of $P_i(t+1)$. This $\left(1-\frac{1}{2^{r_i}}\right)$ can be considered as drop rate like the influence of α in (1). The purpose of using two parameters is to achieve an optimal update speed. By using different parameters, the shift value can be adapted when the probability changes in the context of obtaining the best effect between the drops rate balances. Instead of the lookup table, VVC uses a linearly quantized probability representation and arithmetic operations to update the

probability, which allows adaptation to occur for more situations without the restriction of 64 states in HEVC.

II. More Parameter Probability Estimation

By providing a more accurate probability estimate of the symbol sequence, the compression efficiency of CABAC can be improved. In VVC CABAC, there are two context models and their quantization estimator requires 15 bits for entire probability presentation, then requires 10 and 14 bits to store two hypothetical probability estimates (see Fig.1).

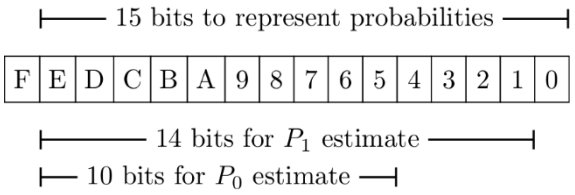

Figure 1. In VVC CABAC, nominal number of 15 bits to represent probabilities, and there are 10 and 14 bits for P_0 and P_1 probability and estimate, respectively.

It means that for each update of P_0 and P_1 within these 15 bits, their high 10 and 14 bits will be kept for probability estimation, but low 1 and 5 bits will be discarded to zero respectively. It implies that $0 < P_0 \leq P_1 < 2^{15}$. Meanwhile, the drop rate α must satisfied with the following,

$$\left(1 - \frac{1}{2^0}\right) < \alpha < \left(1 - \frac{1}{2^{15}}\right) \tag{3}$$

Because the CABAC system in VVC is completely implemented by integers with these divisions being achieved by bit-shift operations, the number of shifts cannot exceed the defined 15 bits. Knowing these conditions, we obtain the range of parameters as $r_i \in \{1,2,\ldots,14\}$ with $r_0 < r_1$. In fact, VVC's CABAC system is equivalent to HEVC's if $r_0 = r_1 = 6$. The probability will be divided by 2^6 that can correspond to the 64-probabilities states in the CABAC system in the HEVC standard. In view of this, we propose more parameters based on these conditions to find a more accurate probability estimation (see Fig.2).

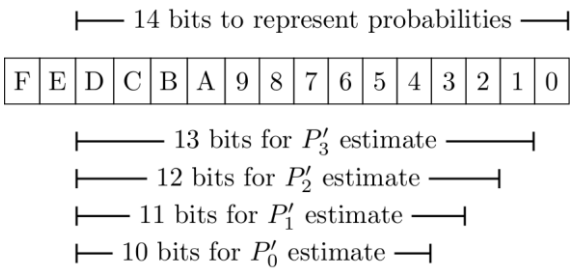

Figure 2. Our proposed, nominal number of 14 bits to represent probabilities, and there are 10, 11, 12 and 13 bits for P_0', P_1', P_2' and P_3' probability and estimate, respectively.

As indicated in Fig.2, our approach will require 14 bits to handle the probability estimation. The reason is that the highest bit is used to represent the MPS symbol, and we also reserve the second high bit to prevent the overflow issue when the updated probability is obtained by $\frac{1}{4}\sum_{i=0}^{3} P_i'$. It also implies that $0 \leq P_0' \leq P_1' \leq P_2' \leq P_3' < 2^{14}$. By the same token, a corresponding drop rate α' in our cases must satisfied with:

$$\left(1 - \frac{1}{2^0}\right) < \alpha' < \left(1 - \frac{1}{2^{14}}\right) \tag{4}$$

and the range of more parameters is $r_i' \in \{1,2,\ldots,13\}$ with $r_0' < r_1' < r_2' < r_3'$.

A. Drop Rate Consideration

Similarly to the approach of VVC, the proposed method applied more adaptation rates for context model, such that they adapt differently to the (MPS) symbol statistics. However, in VVC's CABAC system, the r_i values have been preset for different coding modes so that these drop rates cannot be adjusted at will. Therefore, we tend to keep using these preset parameters (as (5)) for compatibility.

$$r_0' = r_0 \tag{5a}$$

$$r_1' = \frac{3}{4}r_0 + \frac{1}{4}r_1 \tag{5b}$$

$$r_2' = \frac{1}{4}r_0 + \frac{3}{4}r_1 \tag{5c}$$

$$r_3' = r_1 \tag{5d}$$

Considering that using only two parameters in VVC will not be sufficient to allocate more context models for our proposed approach. We will first use (5a) and (5d) to determine their boundaries, then use linear interpolation to obtain the rest of the parameters. The reason is that it also complies with $r_0' < r_1' < r_2' < r_3'$, and can also be easily implemented by bit operations. Our proposed context model, which can replace $P_i(t+1)$ in the probability update of (2) with $P_i'(t+1)$, yields four estimators depending on the drop rate r_i' as follows:

$$P_i'(t+1) = \begin{cases} \left(1 - \frac{1}{2^{r_i'}}\right) \cdot P_i'(t) & MPS \\ \frac{1}{2^{r_i'}} + \left(1 - \frac{1}{2^{r_i'}}\right) \cdot P_i'(t) & LPS \end{cases} \tag{6}$$

B. MPS Determination

According to the definition of CABAC, $P(t+1)$ in (1) represents the probability of receiving LPS, so it must satisfied with $P(t+1) \leq 0.5$. (1) also implies that $P(t+1) < P(t)$ if the received symbol is MPS, otherwise $P(t+1) > P(t)$. Especially, the symbols of MPS and LPS will swap when $P(t+1) > 0.5$, caused by receiving LPS continuously. Thereby, we can determine the symbol of the current MPS by $P(t+1)$. This inference can also be applied to the next generation and is expressed as follows:

$$P'(t+1) = \frac{1}{4}\sum_{i=0}^{3} P_i'(t+1) > 0.5 \quad (7a)$$

$$\rightarrow \sum_{i=0}^{3} P_i'(t+1) > 2.0 \quad (7b)$$

Corresponding to Fig.2, the range of these probabilities from [0.0,1.0] will be scaled to $[2^0, 2^{14}]$ with integers. Therefore, the decision value 2.0 in (7b) will be up-scaled to $2 \times 2^{14} = 2^{15}$, and it is at the highest bit (index F in Fig.1 and 2). That is why we mentioned that the highest bit is used to represent the MPS symbol.

Fig.3 illustrates the probability updating process for a given system (bin value) encoding in the regular mode. Our proposed four estimators will first calculate their probability by (6) and according to the received symbol independently. Next according to (7), we must consider whether to swap the symbol between MPS and LPS if the received symbol is not MPS. Then, the updated probability will be the average value of the estimated probability; this is then passed to the encoding process.

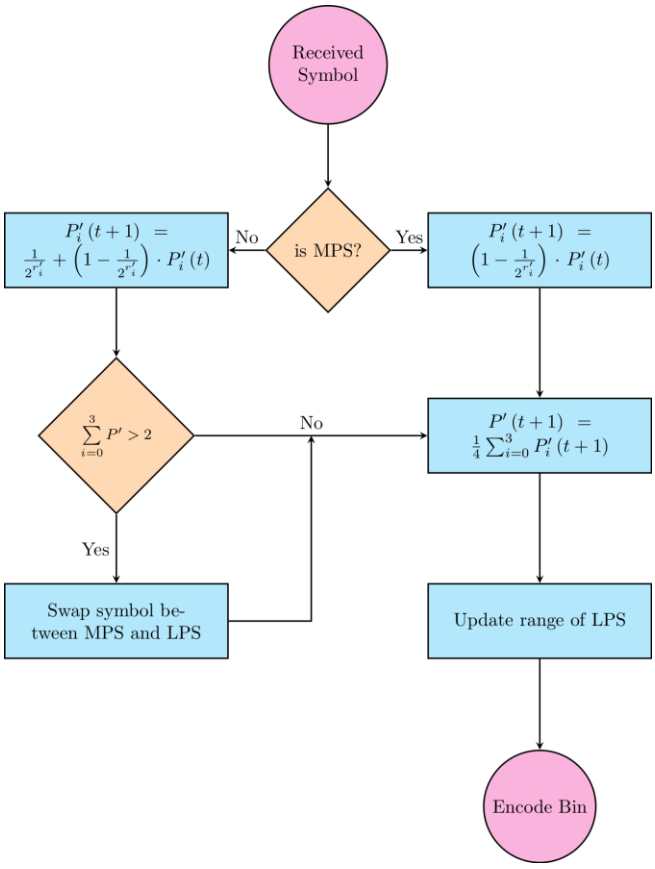

Figure 3. Flow chart for processing the probability and symbol processing between MPS and LPS. The resulting probability will be used to update the LPS range.

In AVC and HEVC, this part is achieved by a set of lookup tables, so the probabilities can only provide a finite number of results, which is poor for floating point approximation if higher precision is required in the near further. Using the linearly quantized approach can be advantageous since the number of digits (in Fig. 1 and 2) of its accuracy can grow at any time. More precisely, it uses a sub-sampling state transition mapping, where different adaptive rates are approximated by bit shift operations. The linear quantization estimator proposed by VVC requires 10 and 14 bits to store the probability estimates of these two hypotheses, depending on the number of probability estimators required for a particular coding application. Compared with the linear quantized version, our proposed method extends it to require 11 to 13 bits to achieve better compression efficiency, and at the same time confirms that these schemes have more expansion in the next generation.

III. Experimental Results

In order to evaluate the gains provided by the proposed checksum validation processes, we encoded the first 64 frames of sequences with different resolutions using the VVC Test Model (VTM) [8] reference profile of the VTM software in version 9.0. All experimental setting conform to the provided VTM test conditions and SDR reference [9]. The Windows 10 operating system and Intel i7-8700K processor at 3.70GHz is used for the simulation platform. For performance evaluation, these sequences are coded in All-intra, Inter-IPPPP and Inter-IBBBP format at a frame rate of 32Hz, and all the sequences are encoded with $QP \in \{20, 24, 28, 32, 36, 40\}$, in which the coder in VVC is used as the benchmark to compute the BD-RATE and BD-PSNR [10]. These experimental results are for GOP: IIIII, IPPPP and IBBBP structures, respectively.

Table I indicates the BD-RATE and BD-PSNR results of proposed method. For sequences with different resolutions, multi-probability estimation will provide more context modeling compatible with reference performance, while providing more prediction solutions. Compared with the reference configuration, they can achieve better coding performance. These experiments show that our method provides higher Luma gain on average compared to the reference results. The proposed BD-RATE gain for context modeling is up to 4.0% for the full frame, and up to 2.0% for the inter-frame mode. To demonstrate the range of BD values evaluated, Figure 4 shows the BD-PSNR and BD-RATE curves for the selected sequences with All-intra mode in each class. These figures clearly show the gain obtained by the proposed method.

As expected, the gain provided with more probability estimators is smaller on average in inter mode. It has little effect on the BD-PSNR of the image in the reference frame. This is because the inter mode only encodes the difference between the motion vector in the reference frame and target to search the best match, resulting in a relatively small bitstream after DCT. Therefore, the improvement in the I frame that will always produce a large number of bits is more obvious than the improvement of P and B frames. It should also be noted that BD-RATE is shown to provide more gain in high-resolution video sequences encoded in all mode. This is because the reduction of bitrate gained by the number of bit shifts depends on the range estimation: using linear quantization can gain a little in range updating.

Therefore, the percentage of bit rate reduction will decrease as resolution increases.

TABLE I. PROPOSED METHODS VS. REFERENCE. APPLYING GOP STRUCTURE AS ALL-INTRA/INTER-IPPPP/INTER-IBBBP

H.266/VVC		All-Intra		Inter-IPPPP		Inter-IBBBP	
Resolution	Sequence	BD-RATE-Y (%)	BD-PSNR-Y (dB)	BD-RATE-Y (%)	BD-PSNR-Y (dB)	BD-RATE-Y (%)	BD-PSNR-Y (dB)
Class A 2560 × 1600	*Traffic*	-2.17	+0.11	-0.21	+0.01	-0.12	+0.01
	PeopleOnStreet	-1.99	+0.10	-0.76	+0.04	-0.69	+0.03
	NebutaFestival	-4.27	+0.30	-0.50	+0.04	-1.50	+0.04
	SteamLocomotive	-3.69	+0.22	-2.29	+0.06	-2.22	+0.06
Class B 1920 × 1080	*Kimono*	-2.46	+0.10	-1.00	+0.04	-0.83	+0.03
	ParkScene	-2.52	+0.11	-0.79	+0.02	-0.72	+0.02
	Cactus	-2.23	+0.08	-0.80	+0.02	-0.70	+0.02
	BQTerrace	-2.26	+0.11	-0.96	+0.02	-0.79	+0.02
	BasketballDrive	-1.64	+0.04	-0.32	+0.01	-0.15	+0.01
Class C 832 × 480	*RaceHorsesC*	-2.36	+0.14	-1.28	+0.05	-1.49	+0.05
	BQMall	-1.02	+0.06	-1.65	+0.07	-2.13	+0.09
	PartyScene	-1.69	+0.11	-0.37	+0.01	-0.30	+0.01
Class D 416 × 240	*RaceHorses*	-0.33	+0.02	-3.39	+0.14	-4.44	+0.18
	BQSquare	-0.81	+0.06	-1.60	+0.07	-1.59	+0.07
	BlowingBubbles	-0.12	+0.01	-2.26	+0.08	-2.28	+0.08
	BasketballPass	-1.42	+0.08	-4.56	+0.20	-5.00	+0.22
Class E 1280 × 720	*FourPeople*	-0.90	+0.05	-0.97	+0.04	-1.26	+0.06
	Johnny	-0.12	+0.01	-2.31	+0.07	-2.39	+0.08
	KristenAndSara	-1.01	+0.05	-1.57	+0.06	-1.45	+0.05
Class F*	*BasketballDrillText*	-0.48	+0.02	-1.02	+0.05	-1.45	+0.06
	ChinaSpeed	-1.43	+0.13	-0.34	+0.02	-0.02	+0.01
	SlideEditing	-1.72	+0.25	-1.14	+0.16	-1.31	+0.19
	SlideShow	-0.41	+0.03	-0.28	+0.02	-0.82	+0.07

*Class F is the non-camera captured content such as video screen content.

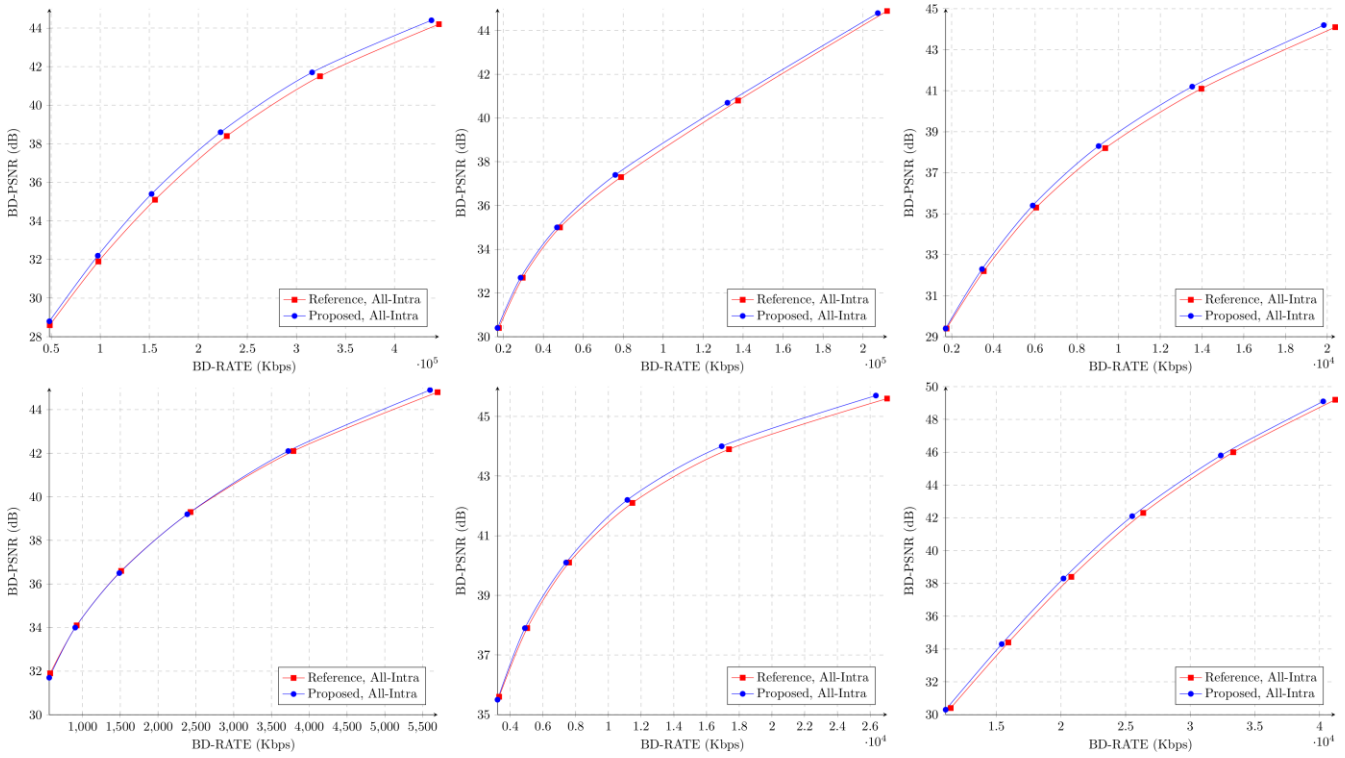

Figure 4. RD curves proposed versus reference with All-Intra modes, from left to right and top to bottom are *NebutaFestival*, *BQTerrace*, *RaceHorsesC*, *BasketballPass*, *KristenAndSara* and *SlideEditing*.

IV. CONCLUSION

In this paper, we use different drop rates to achieve flexibility in order to make the estimators suitable for sources with different statistical information. A major design aspect is the compatibility of having more drop rate decisions and making use of linear quantization instead of finite states in CABAC. This can be achieved by using a more compact representation of the probability state, and a multiplication-free operation can be achieved through a series of bit operations. Experiments show that more estimator representations used for encoding lead to better optimization mode decisions without incurring a significant complexity overhead. The improvement shows that gain is provided in the encoding, and the proposed method can be used in the next generation VVC standard.

REFERENCES

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

[2] B. Bross, J. Chen, S. Liu and Y.-K. Wang, "Versatile Video Coding (Draft 9)," MAY 2020. [Online]. Available: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=10155.

[3] J. Vanne, M. Viitanen, T. D. Hamalainen and A. Hallapuro, "Comparative rate-distortion-complexity analysis of HEVC and AVC video codecs," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22(12), no. IEEE, pp. 1885-1898, 2012.

[4] D. Marpe, H. Schwarz and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H. 264/AVC video compression standard," IEEE Transactions on circuits and systems for video technology, vol. 13(7), no. IEEE, pp. 620-636, 2003.

[5] S.-K. a. C. K.-H. Im, "Higher precision range estimation for context-based adaptive binary arithmetic coding," IET Image Processing, vol. 14(1), no. IET, pp. 125-131, 2020.

[6] E. Belyaev, M. Gilmutdinov and A. Turlikov, "Binary Arithmetic Coding System with Adaptive Probability Estimation by" Virtual Sliding Window"," in 2006 IEEE International Symposium on Consumer Electronics, 2007.

[7] A. Alshin, E. Alshina and J. Park, "High precision probability estimation for CABAC," in 2013 Visual Communications and Image Processing (VCIP), 2013.

[8] J. Chen, Y. Ye and K. S.-H. Kim, "Algorithm description for Versatile Video Coding and Test Model 5 (VTM 5)," 2019.

[9] F. Bossen, J. Boyce, K. Suehring, X. Li and V. Seregin, "JVET common test conditions and software reference configurations for SDR video," 2019.

[10] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves (VCEG-M33)," in VCEG Meeting (ITU-T SG16 Q. 6), 2011.