Scalable HEVC Intra Frame Complexity Control Subject to Quality and Bitrate Constraints

Yuebing Jiang
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, New Mexico 87111
Email: yuebing@unm.edu

Cong Zong
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, New Mexico 87111
Email: czong@unm.edu

Marios Pattichis
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, New Mexico 87111
Email: pattichi@unm.edu

***Abstract*—We introduce an optimal approach for minimizing the computational complexity of HEVC intra encoding subject to constraints in bitrate and reconstruction quality. Our constraint-optimization approach provides an extension to the use of bit constrained rate-distortion optimization so as to minimize encoding time while also delivering sufficient video quality.**

For our approach, we adaptively control the quantization parameter (QP) and a quad-tree-depth oriented coding tree unit configuration to deliver performance that is optimal in the complexity-rate-quality performance space. Our proposed approach employs a spatially adaptive model that uses neighboring configurations to estimate optimal values for QP and the coding tree unit configuration.

We tested our approach using an HEVC standard test video and the ability to dynamically reconfigure between low, medium and high profiles. We found that we can meet the constraints (at 93% (low), 83% (medium) and 93% (high)), while delivering encoding time savings of 13%, 49% and 40% respectively.

***Index Terms*—HEVC, intra coding, coding tree units, optimization**

I. Introduction

HEVC promises significant bitrate improvements over previous video compression standards. Compared to H.264, at the same level of perceptual video quality, HEVC aims to provide a 50% reduction in bitrate [1]. Unfortunately, to achieve a bitrate reduction of 50%, HEVC relies on extensions of the current compression methods that come at significant levels of additional computational complexity. Among the most important extensions to HEVC compression methods, we have recursive coding/transforms units, complex intra prediction modes and asymmetric inter prediction unit division. For intra coding, the current paper proposes the use of a constraint optimization approach that minimizes computational complexity while providing sufficient video quality and compression at the available bitrate.

There has been strong research interest in reducing HEVC encoding complexity for both inter and intra coding. For reducing the computational complexity for inter coding, the authors in [2] introduce the use of different configuration modes. For reducing the computational complexity for intra coding, we note the use of a rough mode set (RMS,[3]), gradient based intra prediction [4], and coding unit (CU) depth control [5]. Unfortunately, these prior approaches did not take into account that video compression requirements can vary with network conditions, energy/power constraints, or varying expectations of video quality. Thus, it is not sufficient to reduce computational complexity without considering the implications on bitrate and video quality. Furthermore, the need for a new approach is particularly important in mobile systems as we discussed in [6], [7].

The current paper reports on a new, efficient implementation of the minimum complexity mode introduced in [6], [7]. To define the minimum complexity mode, let C represent computational complexity as measured by execution time, Q represent a metric of image quality (e.g., PSNR), and R represent the bitrate in bits per sample. Furthermore, to avoid sacrificing video quality, we require that $Q \geq Q_{min}$. Similarly, bitrate constraints are captured by requiring that $R \leq R_{max}$. We also let $\mathcal{C}$ represent the set of all possible configurations. Then, the minimum complexity mode selects the optimal configuration `conf` that solves

$$\min_{\texttt{conf} \in \mathcal{C}} C \text{ subject to: } (Q \geq Q_{min}) \ \& \ (R \leq R_{max}). \quad (1)$$

The constrained optimization formulation of (1) represents an extension of the standard rate control (RC) methods of HEVC. Here, we note that the HEVC reference software implements three methods for intra coding: (1) the unified RQ model [8], (2) the R-lambda model [9], and (3) SATD based rate control [10]. In general, rate-control methods aim to meet the rate constraint (e.g., $R \approx R_{max}$) but do not consider computational complexity. As given in (1), our goal is to extend rate-control research to meet video quality constraints as well as to minimize the required encoding time.

In summary, the contributions of the current paper include: 1) the development of a control mechanism that solves the optimization problem given in (1) for HEVC intra encoding based on the Coding Tree Unit (CTU) level, and 2) the effective implementation of the control mechanism using CTU performance models.

The rest of the paper is organized as follows. In section II, we describe the proposed methodology. The results are given in section III. Concluding remarks are given in section IV.

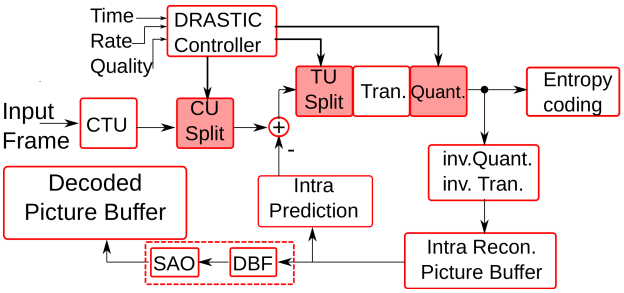

Fig. 1. System diagram for DRASTIC HEVC intra encoding system.

II. Methodology

A. HEVC intra encoding system

We provide a system diagram of the proposed approach in Fig. 1. The most important innovation in our system is the introduction of the controller that is used to handle the optimization process. We use the term *DRASTIC controller* to refer to Dynamically Reconfigurable Architecture System for Time-varying Image Constraints controller as was first introduced in [11]. The DRASTIC controller is provided with measurements of encoding time, rate, and image quality that it uses to select methods for splitting the coding and transform units and to set the quantization parameter for the next incoming frame.

In HM, for Luma prediction, a rough mode set (RMS) [12], [3] that includes 8 modes for 4x4 and 8x8 CU and 3 modes for the remaining CUs. To select the best RMS mode, it estimates the RD performance by using the sum of square differences as an estimate of the distortion and estimate rate based on the number of bits required for the largest transform unit (TU). For chroma prediction modes, the RMS is not needed since there are fewer pixels and modes to work with. The best chroma prediction mode is selected from 5 available modes based on the RD performance as we did for the luma mode. Given the determined best luma and chroma prediction modes, the transform tree and transform coefficients are determined using an exhaustive subdivision process, where the coding tree unit is constructed using splitting as needed so as to provide the best RD performance. The reconstructed pixel values were then saved in the reconstructed picture buffer. To reconstruct the decoded picture, deblocking filter (DBF) and sample adaptive offset (SAO) filtering were supplied. For optimization purposes, we disabled DBF and SAO so as to provide performance measurements that can be controlled directly. Here, we do note that execution times vary significantly for different implementations of the proposed system. However, we expect that the selected prediction modes will remain optimal independent of the specific implementation. In other words, we assume that the same parameters that minimize encoding time for our testing architecture will also produce minimal solutions for different architectures. In future implementations, we certainly expect significant overall time-performance improvements over our current prototype that was constructed using the reference software HM11.0 implementation [13] . Nevertheless, we expect the proposed approach to be directly applicable to faster implementations as well.

B. Optimal configuration management based on scalable parametrization

We specify the optimal configuration based on (1) the quantization parameter, and (2) a scalable parametrization of the CU tree based on `config`. Here, we note that QP also affects encoding time since larger QP values will result in smaller bitrates, lower quality, and lower encoding times since we will have fewer coefficients to encode. On the other hand, `config` is used for controlling the search space for specifying the coding unit sizes.

Refer to Table I and Fig. 2 to see how the `config` parameter is used. The `config` parameter is allowed to vary from 0 to 13. Here, scalability is achieved by making sure that the search space uses a nested subset of the full partition tree. We control the quad-tree partition process using a `process_id` as shown in Fig.2. Here, the `config` parameter gets mapped to a maximum value of the `process_id`. Thus, we do not consider partitioning beyond the maximum value of the `process_id`. For example, for `config` $= 0$, we will not consider any splitting. For `config` $= 1$, the original 64x64 coding unit can be split into 4 32x32 regions, but we will only allow splitting except first 32x32 region. The decision on whether splitting is optimal or not is decided using RD optimization. For `config` $= 6$, we illustrate the search tree using a green line in Fig. 2. Tree space search is performed using depth first search (DFS) (as implemented in HM software). this mechanism is applied to TU control also, unless a split is needed, i.e. there is no 64x64 TU, we will accept split to 32x32 TU.

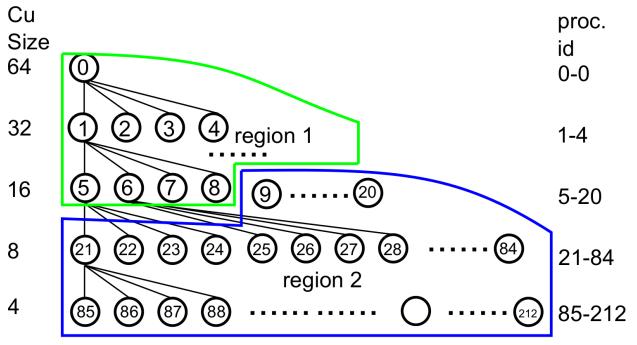

Fig. 2. CU partition control based on the `config` parameter. In this example, we demonstrate the case when `config` $= 6$. In this case, we have $stp\_proc\_id = 9$ which prohibits any splitting for processes with $id >= 9$.

TABLE I
Depth control for CU and TU using `config`. We have `config` $= stp\_proc\_id$ for `config` $\in [0, 5]$. For `config` > 5, we show the relationship below.

`config`	6	7	8	9	10	11	12	13
stp_proc_id	9	13	17	21	37	53	69	85

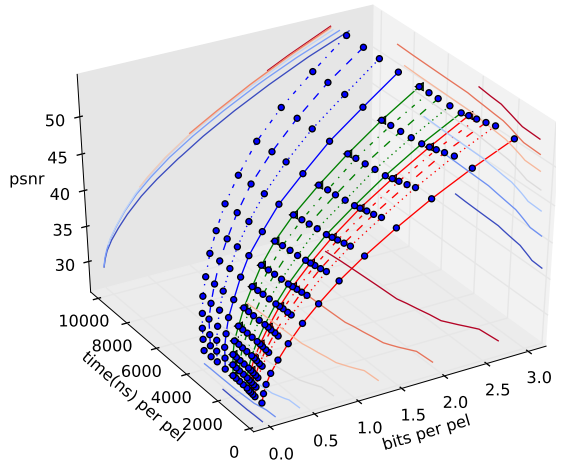

Fig. 3. Rate-distortion-complexity performance example. The graphs were generated by varying QP and `config`. The results were calculated using the median values for the first 6 frames of RaceHorses video (832x480). BPS stands for bits-per-sample. SPS stands for seconds-per-sample. Complexity-distortion-rate performance surface.

The proposed scalable approach can be used to generate a Time-Rate-Quality performance space. We demonstrate this space in Fig. 3. Here, for each plot, we measure (i) time using the number of seconds per sample (SPS), (ii) rate based on the number bits per sample (BPS), and (iii) quality using PSNR (dB). In this example, we use the first 6 frames of the standard RaceHorsesC video (832x480) and produce the median objective surface plot in Fig. 3. To generate the space, we vary QP in the range of $[6, 51)$ with a step of 3 and consider all 14 possible values for `config`. In total, we have 340 possible combinations that have been verified to be optimal in the multi-objective sense (Pareto optimal). As expected, as we increase `config`, we obtain better Rate-Distortion performance at the prize of increased computational complexity. On the other hand, higher values of QP will produce configurations that require lower bitrates with lower quality and reduced computational complexity.

C. Performance prediction model

We also develop an empirical model for fitting the performance surface as a function of QP and `config`. For the model, quality is measured in terms of the mean of square error (MSE), time is measured in ns (10^{-9} second) required for processing a single pixel, and the Rate denotes the number of bits per sample.

We consider an example in Fig. 3 using variations of the model described in [7]. Here, we considered extensions up to 4th order polynomials in QP and `config`. However, to reduce the number of unknown parameters, we only considered models with three coefficients. Using cross-validation, the optimal model was determined to be:

$$\begin{aligned} \text{MSE} &= q_0 \cdot QP^4 + q_1 \cdot \texttt{config} + q_2. \\ \text{Time} &= t_0 \cdot QP + t_1 \cdot \texttt{config}^2 + t_2. \\ \text{Rate} &= r_0 \cdot 2^{-\alpha \cdot (QP - 4/6)} + r_1 \cdot \texttt{config} + r_2. \end{aligned} \tag{2}$$

for $\alpha = 0.56$. In (2), we note that we needed a higher order QP power for the MSE as opposed to the what was needed for fitting the Time objective. Here, we note that our model is dynamical and adjusts to the input sequence. In what follows, we also describe how to update the model based on local measurements.

D. Model update

We note that the linear model described in (2) will need to be updated throughout the video frame. Here, we introduce a linear model update and keep α fixed. For the model update, we use the three neighboring CTUs to update the current CTU as depicted in Fig. 4. As an example, for updating the MSE model, we use:

$$\begin{bmatrix} \hat{q}_0 \\ \hat{q}_1 \\ \hat{q}_2 \end{bmatrix} = \begin{bmatrix} QP_0^4 & \texttt{config}_0 & 1 \\ QP_1^4 & \texttt{config}_1 & 1 \\ QP_2^4 & \texttt{config}_2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \text{MSE}_0 \\ \text{MSE}_1 \\ \text{MSE}_2 \end{bmatrix} \tag{3}$$

which assumes that the neighboring CTUs provide three linearly independent equations $\text{MSE}_i = q_0 \cdot QP_i^4 + q_1 \cdot \texttt{config}_i + q_2$ for $i = 0, 1, 2$. When the linear independence assumption is violated, we select the neighboring CTU coefficients that gave the most accurate predictions. Thus, in our MSE example, instead of estimating q_0, q_1, q_2, we simply select the q_i values used in one of the three neighboring CTUs that gave the most accurate prediction. We apply the same approach for Time and Rate prediction.

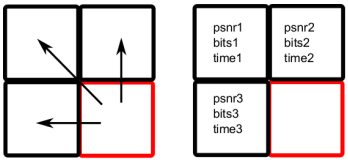

Fig. 4. Model update using 3 neighboring CTUs.

E. Non-linear parameter estimation

Given the fitted model, the goal of the minimum encoding time mode is to determine the optimal configuration. To specify the optimal configuration, we need to determine the optimal integer values of QP and `config` that can produce PSNR that stays above or equal to Q_{min} while the bitrate remains below or equal to R_{max} and also minimize encoding time.

To find the optimal configuration, we will consider the most interesting and important case when the constraints are met with equality: $Q = Q_{min}$ and $R = R_{max}$ in (1). Thus, in this case, the assumption is that we have tight bounds where it takes all of the available bandwidth to reconstruct the video of sufficient quality. Here, it is clear that if this assumption is violated, we cannot expect to be able to satisfy the constraints. After finding the configuration that meets this assumption, we will then search inside the constraint region to reduce encoding time.

TABLE II
RACEHORSESC (832X480) PROFILE SETTINGS (LOW, MEDIUM, AND HIGH) WITH RATE AND QUALITY CONSTRAINTS.

Profile	Rate Const.(Mbps)	Quality Const.(PSNR)
Low	3.184	29.63
Medium	6.111	32.56
High	10.647	37.87

To accomplish this, let MSE_{max} and $Rate_{max}$ correspond to the case when $Q = Q_{min}$ and $R = R_{max}$. Then, substitute MSE_{max} and Rate_{max} into the first two equations of (2), and eliminate `config` to obtain the nonlinear relationship for QP:

$$\begin{aligned} 0 = \quad & (q_1 \cdot \text{Rate}_{max} - r_1 \cdot \text{MSE}_{max}) + (q_r \cdot r_1 - q_1 \cdot r_2) \\ & + (q_0 \cdot r_1 \cdot QP^4 - q_1 \cdot r_0 \cdot 2^{\alpha(4-QP)/6}). \end{aligned} \tag{4}$$

We solve (4) using Newton's algorithm to get a continuous-value for QP (initializing the search with $QP = 20$). We then substitute QP back in (2) to obtain the corresponding value for `config`. This process generates a continuous-valued pair $(QP, \texttt{config})$ that is relaxed to integer values using the ceiling function $(QP_s, \texttt{config}_s) = (\lceil QP \rceil, \lceil \texttt{config} \rceil)$. We then apply a local search in the domain of $[QP_s - 2, QP_s + 1] \times [\texttt{config}_s - 2, \texttt{config}_s + 1]$ to find a feasible solution with the minimum execution time Time.

III. RESULTS

To demonstrate the advantages of the proposed approach, we used a dynamic reconfiguration example. The goal of our example is to demonstrate the ability to switch from a low profile mode to a medium and then a high profile mode. More generally, we refer to [14] for HM-11's rate and quality for difference test videos.

In this example, we define the low, medium, and high profiles by fixing QP to QP=27, 32 and 37 respectively. Furthermore, for comparing to the proposed approach, for controlling both the bitrate and PSNR, we use the full range depth configuration (`config`=13) and reduce the resulting PSNR constraints a little bit to generate the low, medium, and high profiles given in Table II.

We compare the results for the fixed QP configuration shown in Fig. 5 with our proposed approach that is shown in Fig. 6. For constraint satisfaction, we allow mild violations in the order of 10% of the constraints given in Table II. Then, as shown in Fig. 6, we can see that DRASTIC control achieves constraint satisfaction at the high rates of 93% for low, 83% for medium, and 93% for the high profile. Furthermore, compared to the fixed QP results, the proposed approach achieves savings of 13% for the low, 49% for the medium, and 40% for the high profile.

IV. CONCLUSION

In this paper, we presented an optimal approach for minimizing the computational complexity of HEVC intra encoding subject to bitrate and quality constraints. We provide

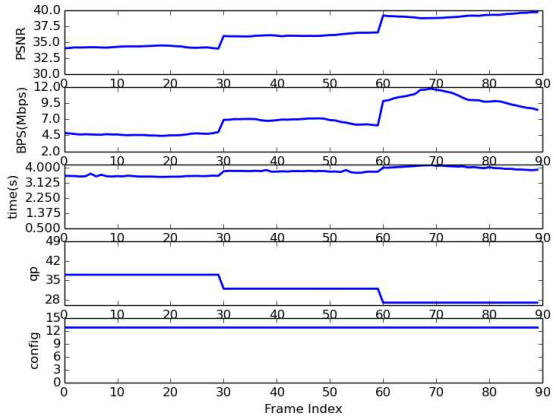

Fig. 5. RaceHorsesC dynamic adaption to switch from low to medium and high profiles using QP only (37,32,27). Refer to Table II for the low, medium, and high profiles.

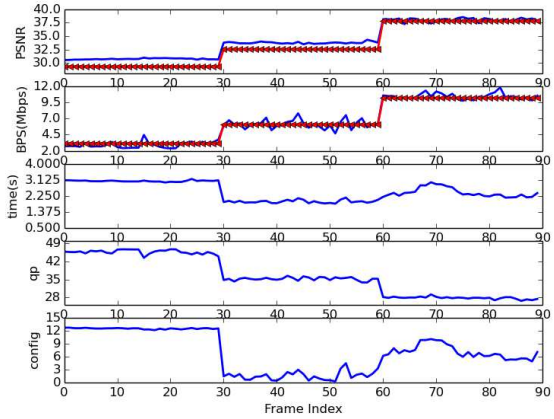

Fig. 6. RaceHorsesC DRASTIC minimum complexity mode. In this example, the proposed approach meets the constraints given in Table II while also minimizing the encoding time.

an effective control mechanism that dynamically adjusts the quantization parameter and the coding tree unit partition mechanism so as to achieve variable constraints on bitrate and video quality. The model is dynamically updated based on the input video. Future work of the proposed model will focus on demonstrating the approach on larger video database and on providing new approaches that can adapt to video content.

V. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under NSF AWD CNS-1422031.

REFERENCES

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

[2] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC complexity and implementation analysis," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 22, no. 12, pp. 1685–1696, 2012.

[3] L. Zhao, L. Zhang, S. Ma, and D. Zhao, "Fast mode decision algorithm for intra prediction in HEVC," in *2011 IEEE Visual Communications and Image Processing (VCIP),*, nov. 2011, pp. 1 –4.

[4] W. Jiang, H. Ma, and Y. Chen, "Gradient based fast mode decision algorithm for intra prediction in hevc," in *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, april 2012, pp. 1836 –1840.

[5] G. Correa, P. Assuncao, L. Agostini, and L. da Silva Cruz, "Complexity control of high efficiency video encoders for power-constrained devices," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1866–1874, 2011.

[6] Y. Jiang and M. S. Pattichis, "A dynamically reconfigurable architecture system for time-varying image constraints (drastic) for motion jpeg," *Journal of Real-Time Image Processing*, pp. 1–17, 2014/10/16.

[7] Y. Jiang, G. Esakki, and M. Pattichis, "Dynamically reconfigurable architecture system for time-varying image constraints (drastic) for hevc intra encoding," in *Asilomar Conference on Signals, Systems and Computers*, Nov 2013, pp. 1112–116.

[8] H. Choi, J. Nam, J. Yoo, D. Sim, and I. Bajic, "JCTVC-H0213,rate control based on unified rq model for hevc," in *ISO/IEC JTC1/SC29 WG11*, San Jose, USA, Feb. 2012.

[9] B. Li, H. Li, L. Li, and J. Zhang, "JCTVC-K0103, rate control by r-lambda model for hevc," in *ISO/IEC JTC1/SC29 WG11*, Shanghai, China, Oct. 2012.

[10] X. W. M. Karczewicz, "JCTVC-M0257, intra frame rate control based on satd," in *ISO/IEC JTC1/SC29 WG11*, Incheon,Korea, Apr.,2013.

[11] Y. Jiang and M. Pattichis, "Dynamically reconfigurable DCT architectures based on bitrate, power, and image quality considerations," in *19th IEEE International Conference on Image Processing (ICIP)*, 2012, pp. 2465–2468.

[12] Y. Piao, J. Min, and J. Chen, "Encoder improvement of unified intra prediction," *JCTVC-C207*, 2010.

[13] I.-K. Kim, K. McCann, K. Sugimoto, B. Bross, and W.-J. Han, "JCTVC-M1002,high efficiency video coding (hevc) test model 11 (HM11) encoder description," in *ISO/IEC JTC1/SC29 WG11*, Incheon,Korea, Apr.,2013.

[14] B. Li, G. J. Sullivan, and J. Xu, "Comparison of compression performance of hevc working draft 5 with avc high profile," in *JCTVC-H0360, JCT-VC Meeting, San Jose (February 2012)*, 2012.