

Low Complexity Rate Distortion Optimization for HEVC

Siwei Ma¹, Shiqi Wang¹, Shanshe Wang², Liang Zhao², Qin Yu¹, Wen Gao¹

¹Institute of Digital Media, Peking University, Beijing, China

²Department of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

Abstract

The emerging High Efficiency Video Coding (HEVC) standard has improved the coding efficiency drastically, and can provide equivalent subjective quality with more than 50% bit rate reduction compared to its predecessor H.264/AVC. As expected, the improvement on coding efficiency is obtained at the expense of more intensive computation complexity. In this paper, based on an overall analysis of computation complexity in HEVC encoder, a low complexity rate distortion optimization (RDO) coding scheme is proposed by reducing the number of available candidates for evaluation in terms of the intra prediction mode decision, reference frame selection and CU splitting. With the proposed scheme, the RDO technique of HEVC can be implemented in a low-complexity way for complexity-constrained encoders. Experimental results demonstrate that, compared with the original HEVC reference encoder implementation, the proposed algorithms can achieve about 30% reduced encoding time on average with ignorable coding performance degradation (0.8%).

I. Introduction

The dramatically increasing of high definition (HD) and beyond-HD (e.g. 4kx2k or 8kx4k) videos are creating stronger demand of high efficiency video coding technology, which are beyond the capabilities of the state-of-the-art video coding standard such as H.264/AVC [1]. Therefore, the ITU-T Video Coding Expert Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG) formed the Joint Collaborative Team on Video Coding (JCT-VC) in Apr. 2010, and work towards conducting the high efficiency video coding (HEVC) project [2]. So far, compared to its predecessor H.264/AVC, HEVC has achieved significant improvement in coding efficiency with more than 50% bit rate saving in terms of perceptual quality [3][4].

As expected, the improvement on coding efficiency of HEVC is obtained at the expense of more intensive computation complexity. In HEVC, an adaptive quadtree structure based on the coding tree unit (CTU) was employed, in which three new concepts named coding unit (CU), prediction unit (PU) and transform unit (TU) [5] are introduced to specify the basic processing unit of coding, prediction and transform. This highly flexible coding structure achieves a leap in coding performance and provides the encoder great flexibility to improve the coding efficiency. For evaluating the compression efficiency of each candidate configuration, the encoder usually employs the Lagrangian multiplier optimization technique [6], which is expressed by

$$\min \{J\} \quad J = D + \lambda \cdot R \quad (1)$$

where D and R are the reconstruction distortion and entropy coding bits of a certain unit. J is the Lagrangian rate-distortion (R-D) cost function to be minimized and λ is the Lagrangian multiplier. The minimization process of the R-D cost is the well-known rate-distortion optimization (RDO). In general, to obtain accurate D and R , for each candidate, the encoder has to perform transform, quantization, entropy coding, inverse quantization, inverse transform, and pixel reconstruction, which makes the R-D cost calculation very time-consuming and brings great burden to the encoder implementation [7]. However, with the rapid developments of the portable devices, the discrepancy between the computationally intensive video codec and the limited computational capability of hardware platforms has become a bottleneck, especially in the real time visual communications. In view of these requirements, some pioneer researches on low complexity implementation of HEVC encoder have been done.

In [8][9], the authors analyzed the implementation complexity and the coding efficiency of these advanced coding tools in HEVC. In [10], the authors proposed a complexity control algorithm for HEVC by adaptively choosing the CU splitting depth. In [11], an early TU decision algorithm for high efficiency video coding is proposed. In [12], coded block flag (*cbf*) is used to terminate PU encoding process, and if the *cbf* of an Inter PU in a CU is zero for luma and chroma except for Inter NxN PU, the next PU encoding process for the current CU will be terminated. Another skip mode early termination algorithm is proposed in [13], and the basic idea is that if skip mode is the locally optimal mode of the current CU depth, skip mode is then considered to be global optimal mode and sub-tree computation process can be skipped. Though these techniques have shown good performance, they only deal with one module in HEVC encoder and the whole structure of the encoder has not been fully considered. In this paper, based on an overall analysis of computation complexity in HEVC encoding process, we proposed a low complexity RDO scheme for HEVC by shrinking the number of intra prediction modes, reference frames and CU splitting patterns, which are among the most computation-consuming modules in HEVC.

The rest of this paper is organized as follows. In Section II, we give an overall analysis of the computation complexity in HEVC Test Model (HM) encoder, including the block partitioning structure as well as inter and intra prediction. In Section III, we propose a low complexity RDO scheme by optimizing the process of intra mode decision, reference frame selection and inter CU splitting decision. Experimental results are provided in Section IV, and Section V concludes the paper.

II. Complexity Analysis of HEVC Encoder

2.1 HEVC encoder computational complexity analysis

HEVC offers more possibilities to split a frame into multiple units and more ways of combining different coding tools and parameters. Though this doesn't have significant impact on the decoder from the complexity aspect, it imposes a heavy computation

burden to the encoder by fully leveraging its capabilities. Table 1 shows the time consumption of the major modules in encoder for *Random Access (RA)* and *Low Delay (LD)* main configurations in HM8.0. The complexity of sub-pixel interpolation is included in the motion estimation process. It is shown that the most time-consuming part is the motion estimation as a result of the multiple reference motion compensation and sub-pixel interpolation etc. For all-intra configuration, the coding complexity mainly comes from the mode decision of all available candidate modes, and it is reported that the full RD search of all available candidate modes will achieve -0.4% coding gain, but bring about three times of the encoding time of the fast mode decision algorithm in HM6.0 [14]. Based on the above considerations, in this section, the relationship between the compression efficiency and the corresponding computational complexity is characterized from the coding structure, intra and inter prediction aspects.

Table 1. Encoding time complexity analysis for each encoder module [%].

Configuration	Motion Estimation		Transform	Quantization	In-loop Filter	Others
	Interpolation	Others				
RA-Main	23.6	40.3	3.2	12.2	0.1	20.6
LD-Main	24.2	45.3	2.5	10.8	0.1	17.1
LD P-Main	17.6	40.7	3.6	15.0	0.1	23.0

2.2 HEVC block partitioning structure

As referred in the introduction, HEVC standard employs a highly flexible quadtree coding structure based on CTU. The CTU size can be selected from 64x64, 32x32 and 16x16 in the encoder and specified in the sequence parameter set (SPS). CTU is employed as the root of the coding tree and each leaf of the quadtree is called CU. Therefore, one CTU can be partitioned into multiple CUs and each CU specifies one coding category: inter CU or intra CU. CU can be further split into one, two or four PUs to specify the prediction information. For efficient prediction, HEVC defines two intra PUs and eight inter PUs to support variable size PU. Suppose the CU size to be $2N \times 2N$, in intra prediction, PART_2Nx2N and PART_NxN PU shapes are defined; and in inter prediction, two square shapes (PART_2Nx2N and PART_NxN), two rectangular shapes (PART_2NxN and PART_Nx2N) and four asymmetric shapes (PART_2NxnU, PART_2NxnD, PART_nLx2N and PART_nRx2N) are supported. The residual block of each CU can be transformed with a quadtree structure, which is usually called residual quadtree (RQT), and the transform is performed on each leaf node of this quadtree.

The encoder selects the best coding tree structure, PU subdivision and RQT configuration through exhaustive execution of RDO, which is a very time consuming process. Generally, the coding complexity of RDO increases monotonically with the depth of the CU partitioning since CU is the root of the PU partition and RQT. However, only limiting the depth of the CU partition is not sufficient to cope with the various local characteristics, as small CUs can be applied to the complex regions which large CUs cannot successfully cover. To address this issue, in Section III, we propose an adaptive

tree-pruning algorithm based on the properties of local prediction residuals.

2.3 Intra prediction

For intra prediction, a significant change in HEVC is the use of more flexible block sizes (4x4 to 64x64) and more intra modes [14]. In total, 33 angular prediction modes as well as planar and DC modes are supported. Due to the increase in the number of intra modes, three most probable modes (MPM) for each PU are specified based on the neighboring PU, which makes the fixed length coding possible for the left 32 intra modes.

The increase of intra modes requires a good mode selection heuristics, since taking all the intra mode of each PU size into the RDO process will impose a great burden to the encoder. Therefore, the original HM reference software employs a fast RDO algorithm by combining the rough mode decision and full RD search together. However, the intra direction information of the neighboring blocks has not been explored in the mode decision process. This observation inspired us to analyze the direction information of the neighboring blocks and make full use of it for intra mode decision, and a new fast RDO scheme for intra mode decision is proposed in section III.

2.4 Inter prediction

For inter coding, similar to H.264/AVC, HEVC employs the block-based motion compensation (MC) with multiple reference pictures, but due to the advanced coding tools adopted, it is more complex compared to H.264/AVC. For example, for sub-pixel interpolation, 8-tap DCT based separable interpolation filter (IF) is employed to generate the luma reference pixels for inter prediction, and for chroma component, 4-tap DCT-IF is applied. Moreover, multiple motion vector predictors derived by the advanced motion vector prediction (AMVP) in HEVC increase the motion search candidates by times. And new coding modes, such as merge mode, also increase the complexity of motion estimation and compensation greatly due to the cross reference of the motion information of the spatial and temporal neighboring PUs.

The combination of the quadtree coding structure and multiple reference pictures lead to the fact that the complexity of motion estimation increases linearly with the number of reference frames. It is generally believed that the correlations between two pictures are stronger when their distance is small, and a straightforward way to reduce the complexity of motion estimation is to remove the long distance reference frames. However, this cannot be successfully applied to all occasions. To achieve high efficiency and low complexity motion estimation, in this paper, we will investigate how to manage the multiple reference pictures in the reference frame set (RFS) based on the spatial and temporal correlations among the adjacent frames and CUs.

III. Low Complexity RDO

Based on the analysis of Section II, in this section, we propose a low complexity RDO

scheme, which consists of three parts: fast intra mode decision, adaptive reference selection and CU splitting early termination. The architecture of our scheme is shown in Fig. 1, and it demonstrates that our algorithms apply to the modules that require intensive computations in the encoder. Specifically, in intra prediction, the direction information of the neighboring blocks is made full use of to speed up intra mode decision; in inter CU decision, the correlation between the energy of prediction residuals and CU splitting are exploited to accelerate the CU splitting termination process; and for reference frame selection, the spatial and temporal correlations among the adjacent frames and CUs are employed to shrink the RFS. The experimental results show that the proposed RDO scheme can achieve high compression efficiency with much lower complexity compared to the HM anchor. The details of the proposed scheme will be introduced in the remained of this section.

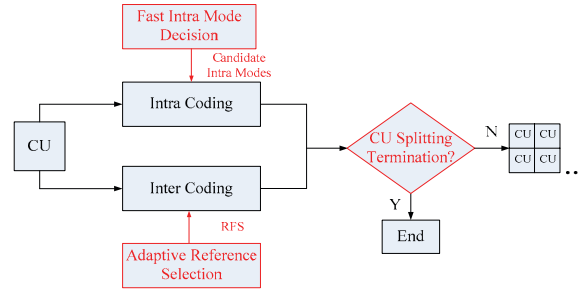


Fig. 1. The architecture of our proposed low complexity RDO scheme.

3.1 Fast intra mode decision

In natural pictures, neighboring blocks usually show similar textures. Consequently, the optimal intra prediction of current block may have strong correlation with its neighboring blocks. Based on this consideration, the MPMs of current block should be always employed as the candidates to compete for the best mode. However, this important information has not been explored in the original intra RDO process of HM. To address this issue, in this subsection, a fast mode selection algorithm for intra prediction is proposed. In our proposed algorithm, we still use the combination of the rough mode decision and RDO process to select the best intra direction. This algorithm has been adopted into the HEVC test model [15]. The architecture of the proposed scheme compared with the default scheme of HM reference software is shown in Fig. 2, and detailed algorithm is described as follows.

Firstly, the candidates selected in rough mode selection are analyzed as an attempt to reduce the number of candidates for various PU sizes. Based on the experiments, we observe that the candidates selected from rough mode decision have a descending trend to be the RDO-optimal mode according to their rank in candidates. Moreover, the first two candidates of all PU sizes have a high probability to be the RDO-optimal mode. This ratio differs a bit with different PU size, so we employ different number of candidates with different PU size for RDO process.

Based on the analysis, we propose our fast intra mode selection scheme. Specifically,

we check whether the MPMs are included in the candidates for each PU size. If they are not included in candidate set, the modes comprised of N best modes from rough mode decision and MPMs will be employed in RDO process. Otherwise, only N best modes will be employed in RDO process. In this work, the number of N best modes in rough mode decision for RDO process is defined to be 8, 8, 3, 3 and 3 corresponding to the PU size of 4x4, 8x8, 16x16, 32x32 and 64x64.

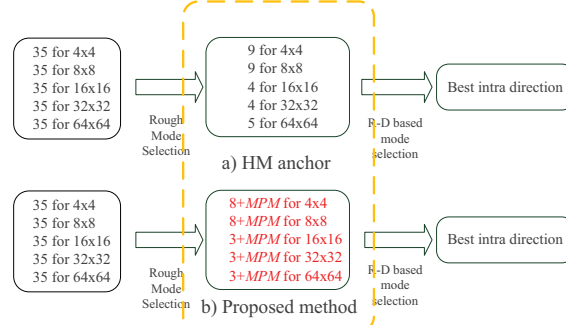


Fig. 2. The architecture of the proposed fast intra RDO scheme compared with the original scheme in HM.

3.2 Adaptive reference frame selection

In this subsection, we propose an adaptive reference frame selection algorithm by shrinking the RFS to reduce the computational complexities of motion estimation. In HEVC inter coding, each frame in RFS has a different probability to be selected as the final reference frame, and the reference frame distribution (RFD) can be employed to investigate the reference frame selection mechanism. Because of the temporal correlations between the adjacent frames, the RFD of the PU in adjacent frames often has great similarity, and this probability is monotonically decreasing as the increase in the temporal distance. Based on this observation, a measurement of motion complexity called C_{ME} is used to determine an initial RFS for a PU, which is defined as follows

$$C_{ME} = \sum_{i=1}^k \alpha_i \times (N_i / \sum_{i=1}^k N_i) \quad (2)$$

where N_i denotes the count number of the i -th reference picture referenced by the encoded PUs of the previous frame and the current frame; k indicates the number of reference frame and is set to be 4 in the paper. α_i is the weighting factor which is determined by the temporal distance of the reference frames.

In this work, we incorporate C_{ME} into the motion estimation process and propose a fast adaptive algorithm for multi-reference frame selection. The algorithm is proposed to adaptively adjust the number of reference frame in the RFS by initial RFS determination and adaptive shrinkage of RFS.

Determination of Initial RFS: For most PUs, it is unnecessary to exhaustively search all of the reference frames. Therefore, in our algorithm, we provide an suitable initial RFS based on C_{ME} to avoid the time consuming motion estimation process in all the reference frames. Specifically, the reference frame whose reference index is outside of the range $[0, C_{ME}]$ is removed from the initial RFS. Moreover, to avoid the converging of

the C_{ME} to 1, we employ the distortion ratio between the current and the next frame as a criterion to expand the RFS. Specifically, if the ratio exceeds a threshold, the RFS is expanded to include the next reference frame.

Adaptive shrinkage of RFS: Generally, most PUs will select the most suitable matching unit in the nearest reference frame, which motivated us to further shrink the RFS to reduce the computational complexity.

Local characteristics of video sequence are exploited to adaptively adjust RFS. Firstly, due to the different size of PU, we define the average distortion per pixel of the PU to be D_p as follows

$$D_p = D_1 / S_{pu} \quad (3)$$

where D_1 refers to the distortion of the PU in the first reference frame and S_{pu} indicates the number of pixels in the corresponding PU. Secondly, in a rate-distortion sense, the determination of the optimal reference frame has great correlation with the motion vector difference (MVD). Therefore, the MVD is employed to facilitate the early searching termination. Based on the above considerations, if the following two conditions in (4) are both satisfied, only the nearest reference frame is included in the RFS

$$D_p < D_0 \quad \text{and} \quad |MVD_x| + |MVD_y| \leq MVD_0 \quad (4)$$

In this paper, the threshold MVD_0 is a constant value and is set to be 1 in this paper. In general, the compression distortion increases monotonically with quantization parameter (QP). Therefore, in our paper, the threshold D_0 is defined by QP as follows.

$$D_0 = c \times \sqrt{QP} \quad (5)$$

where c is a constant and is set as 0.6 in this work.

3.3 Early termination of CU splitting

In video coding, prediction residuals can reflect the prediction accuracy. For temporally stationary and spatially homogeneous blocks, prediction residuals are relatively small, and large CU is more likely to be chosen as the optimal CU size. While small CU partition is preferred for objects with flexible motion since in such cases large CU can be hardly predicted accurately, and thus large prediction residuals need to be coded. Therefore, if the current CU size is sufficient for accurate prediction, there is no need for further splitting. On the contrary, if the residuals of current CU size are large, further splitting might be necessary to get more precise prediction. Based on the above observations, we propose to utilize the energy of the prediction residual block to early terminate the CU recursive splitting process. Suppose the CU size to be $m \times n$, and the energy of the prediction residual block is defined as follows

$$E = \frac{1}{m \times n} \sum_{i=0}^m \sum_{j=0}^n r_{i,j}^2 \quad (6)$$

where $r_{i,j}$ denotes the prediction residuals.

Generally, to early terminate the CU splitting, the energy of the no splitting CUs can be defined as a threshold for classification, but it is not appropriate to be applied here, as a few CUs that didn't split have extremely large energy, and these singular points may significantly enlarge the energy and influence the accuracy of the early termination. Taking this consideration into account, to get a robust threshold value, we employ the k -th smallest energy of no splitting CUs as the threshold as follows

$$E_{thres} = E_k \quad (E_1 < E_2 \dots < E_k \dots) \quad \text{where } k = \text{round}(\beta \times \text{NUM}_{nsi}) \quad (7)$$

The NUM_{nsi} represents the number of CU in the i -th depth that need no splitting and β is an adjustable factor between 0 and 1. By adjusting β , we can change the correctness and error probability of the classification, which plays an important role in balancing the coding efficiency and computational complexity.

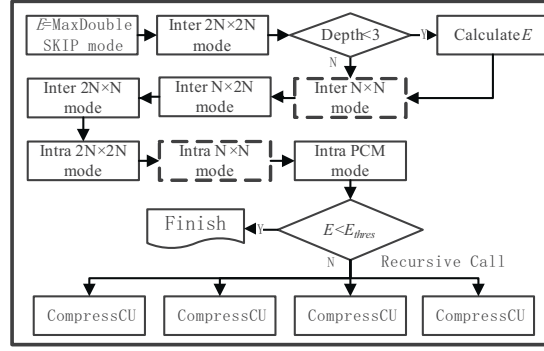


Fig. 3. Architecture of the proposed early termination of CU splitting scheme.

We summarize the whole process of the proposed scheme in Fig. 3. Firstly, the first two inter frames are coded with the conventional mode decision process; and the energy of Inter 2N×2N prediction residual block are obtained from CUs that need no splitting in different depth. For mode decision process of other inter frames, E is initialized with the maximum value. Then SKIP mode and Inter 2N×2N are computed. Based on the inter prediction residual, we calculate the energy of Inter 2N×2N when the current CU size is larger than 8×8, since 8×8 CU cannot be further split into smaller ones. After we sequentially checked other remaining modes in current CU, the energy E is compared to E_{thres} . If it is smaller than E_{thres} , the recursive splitting process is terminated; otherwise, we proceed to the next CU depth.

IV. Experimental Results

To validate the efficiency of the proposed low complexity RDO, we integrate our scheme into the HM8.0 reference software. The experiments were implemented with *Main* setting in *Low Delay* configuration. Experiments were conducted on the common test sequences with quantization parameters 22, 27, 32 and 37 as specified. It is noted that since our intra mode decision scheme [15] is adopted in the HM software, the anchor is generated with the intra coding scheme in [16]. The coding performance is measured by BD-rate and encoder complexity is measured by time saving ΔT , which is calculated as follows

$$\Delta T = (T_{\text{anchor}} - T_{\text{proposed}}) / T_{\text{anchor}} \times 100\% \quad (8)$$

where T_{anchor} and T_{proposed} are the total encoding time of anchor and the proposed encoder, respectively.

Table 2. Performance of the proposed scheme in HM8.0.

	Low Delay (Main)				Low Delay P (Main)			
	Y	U	V	ΔT	Y	U	V	ΔT
Class B	0.9%	1.2%	1.2%	37.13	0.7%	0.9%	0.4%	33.50
Class C	0.3%	0.3%	0.5%	10.80	0.2%	-0.1%	0.3%	9.85
Class D	1.1%	1.9%	1.9%	28.16	0.7%	1.2%	1.4%	23.75
Class E	1.5%	2.3%	2.1%	53.77	1.3%	2.9%	2.3%	49.05
Average	0.9%	1.4%	1.3%	31.43	0.7%	1.1%	1.0%	28.06

In Table 2, we tabulate the RD performance as well as the complexity reduction of the proposed scheme. For the *Low Delay* configuration, it shows that the average PSNR loss compared with the original HM anchor is ignorable (0.8%), while about 30% of the total encoding time on average is saved. The R-D performance comparisons of the proposed algorithm for sequence *RaceHorse* in *Low Delay* are shown in Fig. 4. And it can be seen that the PSNR loss with the proposed algorithm over the full range of QP values is ignorable. The experimental results indicate that our low complexity RDO scheme can efficiently reduce the encoding complexity with practical little performance loss.

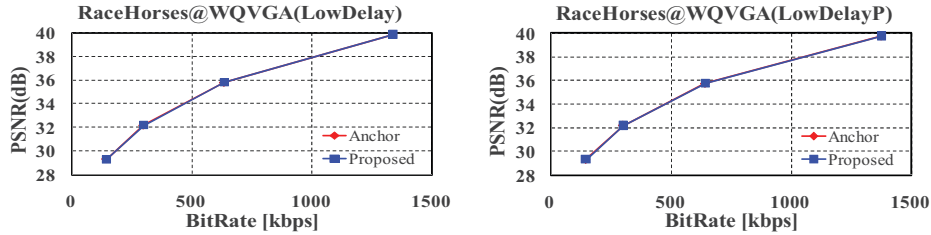


Fig. 4. R-D coding performance of the proposed scheme.

V. Conclusion

This paper proposes a low complexity way to implement the HEVC encoder by accelerating the RDO process. The novelty of this paper lies in that, three techniques which aim to optimize the encoder in a rate distortion sense, are employed to reduce the computational intensive processing in intra mode selection, CU decision and reference frame selection. This paper demonstrates that the computational complexity can be greatly reduced at practically little coding efficiency loss in HEVC. The results and insights of this paper also provide valuable information for the implementation of practical real-time codecs to meet the limitations of power consumption.

Acknowledgement

This work is supported in part by Major State Basic Research Development Program of China (973 Program, 2009CB320903), National Science Foundation of China (61121002, 61103088), National High Technology Research and Development Program of China

(863 Program, 2012AA011505) and National Key Technology R&D Program (2011BAH08B01), which are gratefully acknowledged.

Reference

- [1] ITU-T and ISO/IEC JTC 1, “Advanced Video Coding for generic audio-visual services,” ITU T Rec. H.264 and ISO/IEC 14496-10 (AVC), May 2003.
- [2] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm and T. Wiegand, “High efficiency video coding (HEVC) text specification draft 9 (SoDIS)”, JCTVC-K1003, Dec. 2012.
- [3] G. J. Sullivan, J.-R. Ohm, W. J. Han and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, *IEEE Trans. on Circuits and Systems for Video Technology*, in publication.
- [4] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan and T. Wiegand, “Comparison of the Coding Efficiency of Video Coding Standards–Including High Efficiency Video Coding (HEVC)”, *IEEE Transactions on Circuits and Systems for Video Technology*, in publication.
- [5] I. K. Kim, J. Min, T. Lee, W. J. Han and J. Park, “Block Partitioning Structure in the HEVC Standard”, *IEEE Transactions on Circuits and Systems for Video Technology*, in publication.
- [6] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
- [7] J. Vanne, M. Viitanen, T. D. Hämäläinen and A. Hallapuro, “Comparative Rate-Distortion-Complexity Analysis of HEVC and AVC Video Codecs”, *IEEE Transactions on Circuits and Systems for Video Technology*, in publication.
- [8] G. Correa, P. Assuncao, L. Agostini and L. Cruz, “Performance and Computational Complexity Assessment of High Efficiency Video Encoders”, *IEEE Transactions on Circuits and Systems for Video Technology*, in publication.
- [9] F. Bossen, B. Bross, K. Suhring and D. Flynn, “HEVC Complexity and Implementation analysis”, *IEEE Transactions on Circuits and Systems for Video Technology*, in publication.
- [10] G. Correa, P. Assuncao, L. Agostini and L. Cruz, “Complexity control of high efficiency video encoders for power-constrained devices,” *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 1866-1874, 2011.
- [11] K. Choi and E. S. Jang, “Early TU Decision Method for Fast Video Encoding in High Efficiency Video Coding”, *Electronics Letters*, Vol. 48, pp. 689-691, Jun. 2012.
- [12] R. H. Gweon, Y. L. Lee, J. Lim, “Early termination of CU encoding to reduce HEVC complexity”, JCTVC-F045, Torino, July 2011.
- [13] K. Choi, E. S. Jang, “Coding tree pruning based CU early termination”, JCTVC-F092, Torino, July 2011.
- [14] J. Lainema, F. Bossen, W. J. Han, J. Min and K. Ugur, “Intra Coding of the HEVC Standard”, *IEEE Transactions on Circuits and Systems for Video Technology*, in publication.
- [15] L. Zhao, L. Zhang, X. Zhao, S. Ma, D. Zhao and W. Gao, “Further encoder improvement for intra mode decision”, JCTVC-D283, Daegu, Jan. 2011.
- [16] Y. Piao, J. Min, J. Chen, “Encoder improvement of unified intra prediction,” JCTVC-C207, Guangzhou, Oct. 2010.