MULTIPLE EARLY TERMINATION FOR FAST HEVC CODING OF UHD CONTENT

Ivan Zupancic, Saverio G. Blasi and Ebroul Izquierdo

School of Electronic Engineering and Computer Science, Queen Mary University of London

ABSTRACT

The recently ratified High Efficiency Video Coding (HEVC) standard is significantly outperforming previous video coding standards in terms of compression efficiency. However, this comes at the cost of very high computational complexity, which may limit its real-time usage, particularly when targeting Ultra High Definition (UHD) applications. In this paper, an analysis of HEVC coding on UHD content is presented, showing that on average more than 18% of the total encoding time is spent performing uni-directional Motion Estimation (ME) even when using fast algorithms such as Enhanced Predictive Zonal Search (EPZS). In order to speed up the ME process, a novel approach for fast inter prediction is proposed in this paper based on a Multiple Early Termination (MET) decision process. EPZS is only performed in blocks in which it is needed based on local features of the encoded content, or it is skipped otherwise. Experimental results show that the algorithm achieves on average 9.3% speed-ups over conventional HEVC, at the cost of very small BD-rate losses.

***Index Terms*—** Inter-prediction, video coding, UHD, HEVC, early termination

1. INTRODUCTION

The recently approved H.265/HEVC (High Efficiency Video Coding) standard [1] was developed by the Joint Collaborative Team on Video Coding (JCT-VC), a partnership between the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG), as a successor to H.264/AVC (Advanced Video Coding) [2]. Although HEVC is based on a similar architecture as its predecessor, it comprises several innovative coding tools to maximise the compression efficiency. Some of these novel features are: larger block sizes (referred to as Coding Units, CUs, spanning up to 64×64 samples), recursive CU partitioning (CUs are classified in depths according to the level of recursion), larger frequency transformation block units (up to 32×32 samples), Merge mode in which Motion Vectors (MV) can be inherited from neighbouring prediction blocks, improved MV prediction, better interpolation filters for subpixel Motion Estimation (ME), and so on. Thanks to these improvements, HEVC reportedly achieves 50% bitrate reduction for a given level of objective and subjective visual quality compared to AVC [3]. Unfortunately, this comes at very high computational costs, which limit the real-time usage of HEVC.

These technological breakthroughs in video coding technologies may allow for coding and distribution of content at definitions even higher than High Definition (HD), corresponding to a spatial resolution of 1920×1080 luma samples per frame. In particular, Ultra High Definition (UHD) is a new emerging video format which is being defined with the objective of replacing the HD format in the near future. The specifications of the UHD format are being ratified by the Digital Video Broadcasting (DVB) group and the European Broadcasting Union (EBU) as specified in Recommendation BT.2020 [4], and comprise a spatial resolution of at least 3840×2160 luma samples per frame. Encoding a signal at such resolution represents an extremely demanding compression task, further affecting the computational costs required for HEVC compression. For this reason, methods to decrease the complexity of the HEVC coding scheme are crucial when targeting compression of UHD content.

In this paper, an analysis of typical HEVC coding is presented, which shows that inter-prediction and ME are responsible for a considerable percentage of the encoding time especially when targeting compression of UHD content. A novel algorithm to reduce complexity of HEVC inter-prediction is therefore proposed, based on Multiple Early Termination (MET) of the ME search and downsampling of the distortion metric. The rest of this paper is organised as follows. An overview of related work is presented in Section 2. The analysis of HEVC performance when targeting UHD content is given in Section 3. The proposed algorithm is then detailed in Section 4. Experimental results are presented in Section 5. Finally, Section 6 concludes the paper.

2. RELATED WORK

Several methods to reduce the complexity of HEVC have been proposed. Only methods which target fast HEVC inter-prediction are presented here, as this is the only part of the encoder scheme affected by the proposed approach.

The HM reference software [5], developed by the JCT-VC during the development of HEVC, already makes use of a fast ME algorithm based on Enhanced Predictive Zonal Search (EPZS) [6]. When using EPZS, a certain number of MV candidates are considered, computed using information from spatially or temporally neighbouring blocks, and also testing predefined MVs such as the median MV or the $(0, 0)$ MV. Then, pattern searches are performed in the surrounding of such candidates to find the optimal MV solution for the current block. The algorithm used in the HM reference software is very similar to previous implementations used in the AVC standard; details of this algorithm fall out of the scope of this paper.

Hu and Yang [7] propose a method for reducing the number of MV candidates tested during the motion search based on statistical inference. When using this method, a parameter is computed using the variance of the residual signals in neighbouring blocks. The parameter is then used to decide whether to test a certain MV candidate. Kim *et al.* [8] propose an approach also based on statistical properties of the distortion error during ME, to determine whether to perform bi-directional ME or not. Also, another method was proposed [9] to early determine the usage of the Merge mode on a block.

Email: {i.zupancic, s.blasi, e.izquierdo}@qmul.ac.uk

The authors would like to thank Technology Strategy Board (TSB), UK, for co-funding this work as part of the THIRA consortium project (http://thira.ch.bbc.co.uk/).

This research utilised Queen Mary's MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC grant EP/K000128/1.

Recently, an approach was proposed by Pan *et al.* [10] as a complement to the EPZS algorithm used in the HM reference software, based on Early Termination (ET) of the motion search. In general terms, ET schemes have the goal of reducing the computational complexity of ME by possibly exiting the motion search at an early stage in case a certain condition is triggered. Typically, the ET condition refers to local characteristics of the block, which are assumed to be correlated with the outcome of ME. The method proposed in [10] is applied immediately before performing EPZS. A particular pattern search is performed around the optimal EPZS starting point, and the outcome of such search is used to possibly trigger the ET. In small blocks (namely in blocks extracted CUs at high depths), it is sufficient to examine a small area around the point, which means a small number of MVs is tested in the pattern search. Conversely, a larger pattern search involving testing of more MVs is necessary in larger blocks (extracted from CUs at low depths). Finally, if the original starting point is found to be the minimum solution output after the pattern search, then such starting point is taken as optimal MV for the current block, and the motion search exits without any further calculations. Otherwise, if the minimum solution output after the pattern search is different than the original starting point, EPZS is performed as in conventional HEVC.

3. CHALLENGES OF UHD CONTENT

HEVC complexity becomes an even more critical issue when targeting compression of UHD content. Moreover, existing technology for capturing and recording signals at such definitions is still suboptimal. As a result, typical UHD content available today is characterised by an inherent lack of fine details and other peculiar features. It is reasonable to assume that these features may influence the coding performance, due to their impact on the quality of the reference blocks used for inter-prediction and also on the amount of non-zero residuals resulting after quantisation. These factors affect the complexity and efficiency of HEVC. On the contrary, even though HEVC was designed to address compression of higher resolution signals, very little research exists today to specifically address UHD encoding.

In this section, experiments are presented to highlight some of the challenges while coding UHD content. Tests were performed using the HM reference software version 12.0 under the Common Test Conditions (CTC) specified by the JCT-VC [11] using four different values of the Quantisation Parameter (QP). The encoder was profiled while testing each sequence in order to measure the time spent while encoding CUs at different depths or while performing different tasks. Two groups of test sequences were considered: in the first group (referred to as non-UHD content), 6 sequences at different resolutions spanning from 832×480 to 1920×1080 luma samples were considered. In the second group, sequences at UHD resolution of 3840×2160 luma samples were considered. Finally, the average encoding times were computed for all sequences in each group, at all QPs.

The chart in Figure 1 (a) shows the distribution of encoding time for these two groups broken down with respect to CU depths. This was computed as the average time spent by the encoder testing CUs at a certain depth, with respect to the total coding time. Clearly, while non-UHD content is characterised by a more uniform distribution of times, encoding time is mostly spent at higher depths when coding UHD content. Around 36% of encoding time was spent coding CUs at depth 3 (i.e. 8×8 luma samples) in UHD content, compared with only 28% in the case of non-UHD signals. The chart in Figure 1 (b) shows instead the distribution of encoding time broken down with respect to specific tasks such as EPZS, BiPrediction or Sub-pel ME. Again, UHD content presents a distinct behaviour:

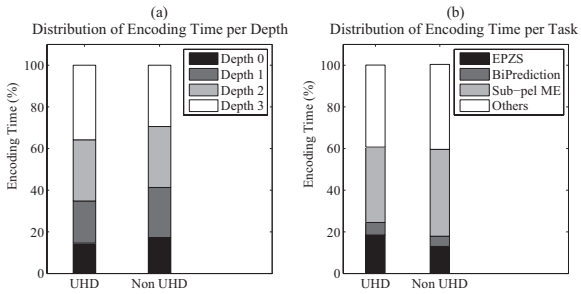

Fig. 1. Encoding time distribution per depth (a) or per task (b) for UHD or non-UHD content.

EPZS takes a consistently higher amount of the total encoding time (18%) than in non-UHD content (less than 13%). Due to such a higher impact of EPZS on the total encoding time, it is crucial that the motion search is as fast as possible when targeting UHD coding.

4. MULTIPLE EARLY TERMINATION WITH SAD SUBSAMPLING

While EPZS is already considerably less complex than full search ME, it still involves testing of a relatively large number of MVs. Each MV is typically examined in a Rate-Distortion (RD) sense. The distortion between original block and prediction block obtained by means of such MV is computed. In conventional HEVC, this happens by means of the Sum of Absolute Differences (SAD) metric. Formally denote as $X[h, w]$ and $P[h, w]$ the samples in the original block X and prediction block P respectively, where $h = 0, ..., H-1$, $w = 0, ..., W-1$ and where H and W are the blocks height and width respectively. Then the SAD is computed as:

$$SAD = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} |X[h, w] - P[h, w]|, \quad (1)$$

The distortion is then adjusted with an estimate of the rate necessary to encode the corresponding MV to obtain an estimated RD cost C. This is typically computed as:

$$C = SAD + \lambda R_{MV}, \quad (2)$$

where λ is a predefined Lagrangian multiplier, and R_{MV} is defined as:

$$R_{MV} = bX + bY \quad (3)$$

where bX and bY are the number of bits necessary to encode the x and y MV component respectively.

Each SAD computation and the corresponding evaluation of the RD cost are computationally expensive for the encoder. Moreover, in many blocks, particularly in static areas of content or slow-motion sequences, the motion search starting point is already capable of providing sufficient prediction accuracy, which means EPZS examines a large number of MVs while finally returning the motion search starting point as optimal solution. All RD cost computations performed during the motion search in these cases do not impact the coding efficiency, while drastically affecting complexity.

The proposed method has the goal of identifying these cases to selectively avoid performing EPZS and consequently reducing complexity. As opposite to other methods which operate directly before or during EPZS (such as the approaches in [7] or [10]), the proposed algorithm acts in the previous step, during the selection of the optimal starting point. The idea is that an ET scheme can be applied

to each of these candidate starting points to check whether any of these may be used as final MV solutions without performing further computations.

In particular, the EPZS algorithm used in HM initially considers a set of N possible candidate MV starting points, referred to in this paper as MV_0, ..., MV_{N-1}. The first candidate MV starting point is derived as the (component-wise) median of the MVs that are used during the MV prediction process; refer to such candidate as MV_0. The second candidate MV starting point is the zero-valued MV (i.e. a MV whose components are both 0); refer to such candidate as MV_1. Finally, up to 3 other candidate MV starting points are considered, namely the optimal MVs previously found in spatially adjacent blocks (left, top, and top-right). Note that not all of these MVs may be available (for instance if any of the neighbouring CUs are intra-predicted, or when currently encoding a block at the borders of the frame). The available MVs are included in the list of possible candidate starting points, referred to in this paper as MV_2, MV_3 and MV_4. In total, a minimum of 2 and maximum of 5 candidates may be considered. In conventional HEVC, each of these points is tested in an RD sense, and the best point is then used as a starting point for EPZS.

In the proposed method, the following algorithm is instead applied while considering each starting candidate MV_i:

1. Compute the estimated RD cost C_i between original block and prediction obtained using MV_i.
2. Perform a diamond pattern search in the surrounding of MV_i to test up to 4 MVs. Let x_i and y_i be the horizontal and vertical components of MV_i respectively. Denote as $MV_{i,0} = MV_i$, and initialise an index j to 1. Then:
   (a) Compute the RD cost $C_{i,j}$ between original and prediction block extracted using $MV_{i,j}$, where
   $$MV_{i,j} = (x_{i,j}, y_{i,j}),$$
   $$x_{i,1} = (x_i - 1), x_{i,2} = (x_i + 1), x_{i,3} = x_i, x_{i,4} = x_i,$$
   and
   $$y_{i,1} = y_i, y_{i,2} = y_i, y_{i,3} = (y_i - 1), y_{i,4} = (y_i + 1).$$
   (b) If $C_{i,j} < C_i$, namely if there exists an $MV_{i,j}$ with $j \geq 1$ whose corresponding RD cost is less than the cost of the original MV starting candidate MV_i, then update the temporary minimum cost to $C_{min} = C_{i,j}$. Consider an enhanced optimal starting point as $MV_{min} = MV_{i,j}$. If $i < N - 1$, increment the index i and to Step 1, otherwise go to Step 4.
   (c) Else if $C_{i,j} \geq C_i$ and $j < 4$, increment j and go to Step 2a. Else, if $j = 4$, the depth d of the current CU is considered. In case $d = 2$ or $d = 3$ (namely the CU is smaller or equal than 16×16 luma samples), go to Step 3, otherwise go to Step 2d.
   (d) Only for blocks extracted from CUs at depth $d = 0$ or $d = 1$ (namely whose size is larger or equal than 32×32 luma samples), an additional hexagonal search is performed after the diamond search. For each value of j such that $4 < j \leq 10$, compute the cost $C_{i,j}$ between original and prediction block extracted using $MV_{i,j}$, where $MV_{i,j} = (x_{i,j}, y_{i,j})$,
   $$x_{i,5} = (x_i - 2), x_{i,6} = (x_i - 1), x_{i,7} = (x_i - 1),$$
   $$x_{i,8} = (x_i + 1), x_{i,9} = (x_i + 1), x_{i,10} = (x_i + 2),$$
   and
   $$y_{i,5} = y_i, y_{i,6} = (y_i + 2), y_{i,7} = (y_i - 2),$$
   $$y_{i,8} = (y_i + 2), y_{i,9} = (y_i - 2), y_{i,10} = y_i.$$
   (e) If $C_{i,j} < C_i$ then update $C_{min} = C_{i,j}$. If $i < N - 1$, increment the index i and go back to Step 1, otherwise go to Step 4..
   (f) Else if $C_{i,j} \geq C_i$ and $j < 10$, increment j and go to Step 2d, otherwise, if $j = 10$ go to Step 3.
3. If $C_i \leq C_{i,j}\ \forall j \geq 1$, and $C_i < C_{min}$, then ET is triggered. MV_i is selected as the MV for the current block, and the algorithm exits.
4. Else, if ET is not triggered. MV_{min} is selected as the enhanced optimal starting point for the motion search, and EPZS performs as in conventional HEVC.

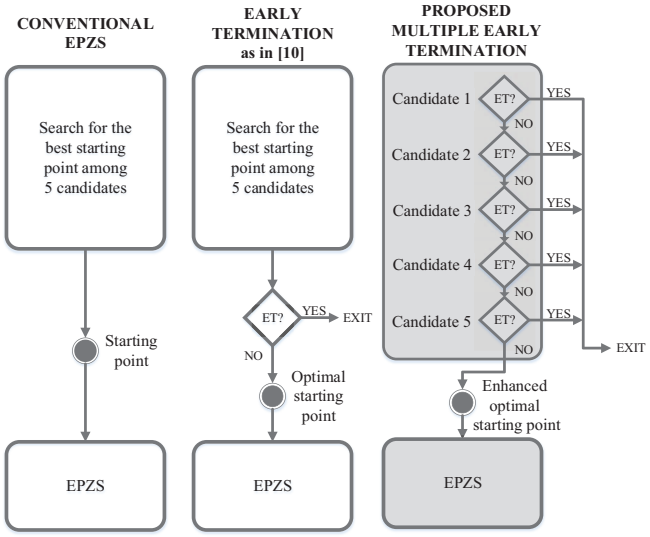

Fig. 2. Block diagram of proposed algorithm compared with conventional HEVC and previously proposed ET algorithm in [10].

A simplified block diagram of the proposed algorithm compared with conventional HEVC and the algorithm proposed in [10] is shown in Figure 2. Notice that the pattern search used in Step 2a is the same as used in [10]. On the other hand, notice also that if ET is not triggered in any of the considered ET points, the algorithm outputs an enhanced starting point MV_{min}, which is the outcome of multiple pattern searches. This is potentially a better starting MV than the one used in conventional HEVC or in [10].

MET involves considerably more SAD computations during the selection of the optimal starting point compared with conventional HEVC or the algorithm in [10]. Such computations may impact the effectiveness of the algorithm illustrated in this section. For this reason, in order to further speed up the MET process, each SAD computation on blocks whose size is larger than 8×8 samples is calculated by means of downsampling.

In particular, downsampling is performed in both dimensions (width and height), in such a way that the downsampled block size is reduced to 8×8 samples regardless of the blocks original size. Formally, the following downsampled blocks, namely $\hat{X}[m, n]$ and $\hat{P}[m, n]$ are considered:

$$\hat{X}[m, n] = X[mM, nN], \quad m, n = 0, ..., 7, \tag{4}$$

and:

$$\hat{P}[m,n] = P[mM, nN], \quad m,n = 0,...,7, \tag{5}$$

where $M = \frac{W}{8}$ and $N = \frac{H}{8}$ are the downsampling rates for each dimension.

Correspondingly, the downsampled SAD is computed as

$$\hat{SAD} = K \sum_{m=0}^{7} \sum_{n=0}^{7} |\hat{X}[m,n] - \hat{P}[m,n]|, \tag{6}$$

where $K = \frac{1}{MN}$ is the scaling factor. The RD cost is then computed using such downsampled SAD as in Eq. 2.

5. RESULTS

The proposed algorithm was tested on 7 UHD sequences. All the tested sequences have a spatial resolution of 3840×2160 luma samples, 8 bits per sample bit depth, and 4:2:0 chroma subsampling. Details for the tested sequences such as framerate, number of frames, Spatial Index (SI), and Temporal Index (TI) [12], can be found in Table 1. Sequences *ManAndPlants*, *ParkAndBuildings*, and *Vehicles* are part of the BBC UHD dataset [13], while sequences *CampfireParty*, *Marathon*, *Runners*, and *RushHour* are part of SJTU 4K Video Sequences dataset [14]. Each sequence was encoded at four different QP values (22, 27, 32, 37), as specified in [11].

Table 1. UHD sequences used for testing

Sequence name	Frame rate	Number of frames	SI	TI
ManAndPlants	50	500	4.52	17.69
ParkAndBuildings	50	500	15.21	41.98
Vehicles	50	500	10.06	19.53
CampfireParty	30	300	5.51	35.52
Marathon	30	300	5.11	19.05
Runners	30	300	6.43	25.35
RushHour	30	300	3.95	19.90

The sequences from Table 1 were tested using the following configuration: Dual 6-core Intel Westmere (E5645) with 24G RAM, using the JCT-VC CTC with Random-Access Main configuration, [11]. For all the tested sequences, the Bjontegaard Distortion (BD) rate [15] was computed. This is a measure of the performance of a proposed encoder with respect to an anchor, in percentage. Positive values of the BD-rate mean a decrease in compression efficiency with respect to the anchor. In addition, the total encoding speed-up was computed for each of the 4 tested QPs for each video using the following formula:

$$\Delta T_i = \frac{T_A - T_M}{T_A} \times 100\%, \tag{7}$$

where T_A denotes the total encoding time for the anchor encoder, and T_M denotes the total encoding time for the modified encoder. Finally, arithmetic mean of ΔT_i for all 4 points was computed to obtain the average encoding speed-up ΔT.

The results can be found in Table 2, where BD_{MET} and ΔT_{MET} denote the BD rate and total encoding speed-up obtained using the proposed method. These were compared with the results obtained using the method in [10], where the BD rate and total encoding speed-up obtained using such method are also presented in the Table 2, denoted as BD_{ET} and ΔT_{ET} respectively.

Table 2 shows that the proposed method consistently decreases HEVC complexity, achieving more than 2.5 times speed-up compared to the method in [10] while keeping the BD rate losses at very low values. As can be seen in Table 1, the results are obtained on a set of sequences with a wide range of SI and TI. In particular, SI is in the range from 3.95 to 15.21, while TI spans from 17.69 to 41.98. On average, the speed-up for the proposed method is 9.3%, compared to 3.7% for the method described in [10], while the BD rate losses are 0.29% and 0.10% respectively.

Table 2. Experimental results

Sequence name	BD_{ET}	BD_{MET}	ΔT_{ET}	ΔT_{MET}
ManAndPlants	0.18%	0.51%	4.3%	10.4%
ParkAndBuildings	0.11%	0.36%	2.4%	7.7%
Vehicles	0.05%	0.11%	3.1%	6.5%
CampfireParty	0.17%	0.35%	4.8%	12.6%
Marathon	0.10%	0.29%	3.7%	9.5%
Runners	0.03%	0.12%	3.9%	8.9%
RushHour	0.07%	0.30%	3.6%	9.6%
Average	**0.10%**	**0.29%**	**3.7%**	**9.3%**

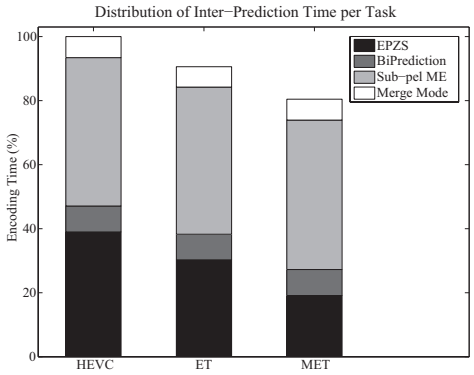

Fig. 3. Inter-prediction time distribution per task for HEVC, ET [10], and proposed MET, in the RushHour sequence.

Speed-ups are obtained as a result of decreasing the impact of EPZS on inter-prediction. Figure 3 shows the distribution of encoding time needed during inter-prediction for particular tasks such as EPZS, BiPrediction, Sub-pel ME and Merge prediction, in the case of conventional HEVC, the method described in [10], and the proposed MET approach, for the *RushHour* sequence. The impact of EPZS on the total inter-prediction time decreases from 38% in conventional HEVC and 33% using the method in [10], to only 23% using the proposed approach.

6. CONCLUSIONS

The HEVC standard significantly outperforms its predecessor AVC, obtaining on average more than 50% higher coding efficiency. This comes at the cost of very high computational complexity, particularly when targeting content at high spatial and temporal resolution, as is the case for UHD applications. In this paper, a novel method to reduce the complexity of HEVC inter-prediction was proposed, based on MET schemes applied to the EPZS search at the same time while selecting the optimal starting point. Also, to further decrease the time needed for the encoding, distortion is measured on downsampled blocks to reduce complexity of SAD operations. The approach was shown achieving on average 9.3% of total encoding time reductions, at a very small cost in terms of compression efficiency degradation, obtaining on average 0.29% BD rate increases with respect to conventional HEVC.

7. REFERENCES

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

[2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.

[3] J. Ohm, G. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards - including high efficiency video coding (HEVC)," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1669–1684, 2012.

[4] ITU-T, "BT.2020: Parameter values for ultra-high definition television systems for production and international programme exchange," ITU-T, Tech. Rep., 2012.

[5] ITU. HM Reference Software. [Online]. Available: https://hevc.hhi.fraunhofer.de/HM-doc/

[6] A. Tourapis, O. Au, and M.-L. Liou, "Highly efficient predictive zonal algorithms for fast block-matching motion estimation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, no. 10, pp. 934–947, Oct 2002.

[7] N. Hu and E.-H. Yang, "Fast motion estimation based on confidence interval," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 8, pp. 1310–1322, Aug 2014.

[8] J. Kim, D. Jun, S. Jeong, S. Cho, J. Kim, and C. Ahn, "An SAD-based selective bi-prediction method for fast motion estimation in high efficiency video coding," *ETRI Journal*, vol. 34, no. 5, pp. 753–758, 2012.

[9] M. Kim, H.-J. Lee, and N. Ling, "Fast merge mode decision for diamond search in high efficiency video coding," in *Visual Communications and Image Processing (VCIP), 2013*, Nov 2013, pp. 1–6.

[10] Z. Pan, Y. Zhang, S. Kwong, X. Wang, and L. Xu, "Early termination for tzsearch in HEVC motion estimation," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 1389–1393.

[11] F. Bossen, "Common test conditions and software reference configurations," JCT-VC, Tech. Rep., October 2012.

[12] ITU-R, "Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications," ITU-T, Tech. Rep., 1999.

[13] R. Weerakkody, M. Naccari, and M. Mrak, "UHD test sequences," JCT-VC, Tech. Rep., November 2013.

[14] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, "The SJTU 4K video sequence dataset," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, July 2013, pp. 34–35.

[15] G. Bjontegaard, "Improvements of the BD-PSNR model," ITU-T SG16/Q6, 35th VCEG Meeting, Doc.VCEG-AI11, 2008.