Pixel Decimation of RD-Cost Functions in the HEVC Encoder

Ahmed M. Hamza; University of Portsmouth
Abdelrahman Abdelazim; American University of the Middle East
Djamel Ait-Boudaoud; University of Portsmouth

Abstract

We present and analyse schemes for the improvement of computational complexity in the current HEVC (High Efficiency Video Coding) standard, by a subsampling of the block-matching distortion cost functions used in the encoding process. HEVC improves on prior standards considerably in coding (compression) efficiency, with a large set-back in time complexity for inter and intra prediction processes and mode decisions. We alleviate this by reducing the number of calculations per decision in all modes of prediction, through pixel decimation in the SAD and SSE distortion cost functions. Experimentation with different patterns shows significant encoding time reduction with these schemes, used in tandem with built-in Fast Encoding optimizations in the HEVC reference implementation.

Introduction

The current H.265/HEVC (High Efficiency Video Coding) standard, developed jointly by ITU-T/Video Coding Experts Group and the ISO-IEC/Moving Picture Experts Group (MPEG), features highly-evolved quadtree based block partitioning structures and coding tools designed, primarily, to achieve enhanced coding efficiency relative to earlier standards. Coding efficiency here refers to the minimization of bit rate for compressed video at a given quality level. HEVC achieves a desired goal of roughly 50% bit-rate improvement [**?**] at the cost of many times the encoding time of earlier encoders. The inevitable increase in complexity, coupled with increases in video resolution and frame rates poses a problem of time and computational power requirements, especially for the now ubiquitous mobile devices and applications of real-time video encoding.

This paper addresses the encoding time complexity by targeting the rate-distortion (RD) optimization process, based on cost functions used by the encoder to make its parameter derivation decisions for the quadtree structure. Most of the work on the subject focuses on eliminating/skipping possibile scenarios (modes) in the quadtree partitioning process [**?**], [**?**]. We instead reduce the load of the block-matching computation of RD cost functions, while minimizing the impact to their decision-making.

Preliminaries

HEVC Block Coding Structure

The general scheme of hybrid block-based video coding in HEVC derives from previous standards, but adds considerable fexibility through several adaptive (partitioning, prediction, transform) mechanisms to help achieve the high efficiency of data representation exhibited by HEVC. The primary block structures are Coding Tree Units (CTU) – analogous to the conventional macroblock – which are subdivided into a number of possible arrangements of Coding Units (CU), the CU blocks varying from 8 × 8 to 64 × 64 in size. These are themselves representable by Prediction Unit (PU) blocks of varying symmetric and assymetric shapes down to 4 × 4 in size. Several overview documents (see in particular [**?**], [**?**], [**?**] and [**?**]) detail a complete description of the high-level syntax and design of the standard.

We are interested in the decision-making parameter derivation processes that consume most of the compute time for an encoder, comprising a large part of the complexity increase in both encoding modes for this standard. A frame can be coded in either intra or inter modes of prediction. Inter-mode coding uses spatial and temporal redundancy of information, i.e., using data from previous frames, and incorporating several modes of operation (merge, skip, etc.,). Intra mode is usually reserved for the initial frame in a Group of Pictures (GOP), using fixed, predefined planar/angular modes in the same frame, as candidate references for a CU encoding.

Inter mode achieves its high rate of compression, while maintaining objective quality, by utilization of spatial and temporal redundancy and an elaborate block-based, quarter-PEL granular motion estimation process. For the reference implementation HM software[**?**], almost all of the above is customizable by configuration, including depth level of the coding trees, search spaces, optimizations and other tools.

In either mode, the best possible candidate(s) determine the CU mode decisions and underlying PU parameters; *best* being chosen through rata-distortion-optimization (RDO) with lagrangian cost functions detailed below. These come in three flavors, as deemed suitable for each type of decision.

Rate-Distortion Optimization Measures

All cost functions J in use are based on a distortion measure D and a bit-cost B, associated with each possible case. The HEVC documentation [**?**] divides the cost functions into three distinct categories:

$$J_{mode} = (D_{SSE,luma} + w_{choma} * D_{SSE,chroma}) + \lambda \times B_{mode} \quad (1)$$

where J_{mode} is used for final candidate selection in intra mode, and mode-selection decisions for inter mode (incl. skip/merge type decisions). Here w_{choma} (weighting factor) and λ are QP based: $\lambda = \alpha W_k 2^{(QP-12)/3}$, where α is set to 1.0 or a lesser value pending number of reference pictures, and with additional adaptive factor W_k for QP offset level.

In prediction-type decisions the form:

$$J_{pred,SAD} = D_{SAD} + \lambda \times B_{pred} \quad (2)$$

is used at integer-pixel levels of granularity, with an equivalent, Hadamard-transformed cost in

$$J_{pred,SATD} = D_{SATD} + \lambda \times B_{pred} \tag{3}$$

used for merge-candidate selection and sub-pixel level motion estimation decisions in inter mode, and for rough candidate selection among the 34 angular and planar possibilities in intra mode. Now we examine the distortion functions D_{SSE} and D_{SAD} which we intend to optimize in this work. They are, respectively:

$$D_{SSE} = \sum_i \sum_j \text{Diff}_{i,j}(\mathbf{C},\mathbf{O})^2 \tag{4}$$

and the similar but more computationally economical

$$D_{SAD} = \sum_i \sum_j |\text{Diff}_{i,j}(\mathbf{C},\mathbf{O})| \tag{5}$$

for an original block **O** being matched against a candidate block **C**, through all rows i and columns j of said blocks. The sum of squared errors (part of, hence correlated to, the Peak Signal to Noise Ratio measure) impacts all modes of operation as noted above, whereas D_{SAD} affects inter (but not intra) coding.

Pixel Decimation Patterns

The elaborate process of arriving at best candidates for the entire Residual Quad Tree structure of a CTU contains many degrees of freedom in mode and candidate selection, particularly with the motion estimation process in inter mode, leading to a combinatorial increase in the number of evaluations through block-matching cost functions (as compared to less sophisticated algorithms of prior standards since H.261 up to H.264/AVC). Reducing the calculations performed within the cost functions therefore positively affects the computational complexity of the encoder. Where subsampling is involved, this comes at the expense of optimal decision making. We investigate here the extent of this expense in coding efficiency and objective quality against time saving gains for the encoder.

Notes on Existing Methods

The HM software comes with by-default activated optimization tool in this regard. Fast Encoding (FEN) set to 1 allows for a skipping of entire rows in the D_{SAD} and D_{SSE} functions for block sizes greater than 8 PEL width.

Other limitations on complexity are also configurable and in effect, by default, for the inter mode motion estimation. A search range (typically 64 PEL) around an originating block is enforced rather than allowing a full search in the space of the frame. Also, the full-search mechanism is replaced by a Test Zone Search (TZSearch) algorithm [**?**] building on search optimizations like Three Step Search and diamond/square patterns. This reduces the number of block-matching operations and candidates for best motion vector, but not the number of operations per block match. FEN also limits the refinement iterations in the TZSearch operation.

Early work by Liu and Zacarrin [**?**] introduced several subsampling methods including pixel decimation, prior to the current set of video coding standards. We base our work on these methods and investigate improved pattern variations, both in tandem and without FEN optimizations in HEVC.

Proposed Schemes

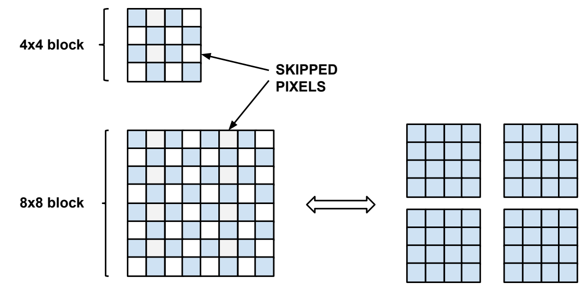


__Figure 1.__ Pixel Decimation in 8×8 block and equivalent non-decimated 4×4 blocks.

Figure **??** shows the basic checkerboard pattern without FEN, yielding a 2-to-1 subsampling ratio. Applying this to individual 4×4 blocks comprises a large information loss for PUs at that size, so we restrict the decimation to everything but the finest level of granularity in block sizes, i.e., widths 8 and up (FEN actually restricts further to blocks of width *greater* than 8 pixels).

Since decimation reduces the number of comparisons by an order of 2, then to compare the distortion cost of the decimated pattern (call this D'_{SAD}) to a normal D_{SAD} for an equivalent block size we multiply the decimated cost by the same subsampling factor to obtain a fair estimate of the non-decimated cost. Thus in general:

$$D_{SAD} = k \times D'_{SAD} \tag{6}$$

where in this example k is 2. FEN accordingly corrects its cost estimate, in the reference code. For our patterns designed on top of the FEN scheme (see **??**) we modify the estimate to account for the increased subsampling factor, 4-to-1.

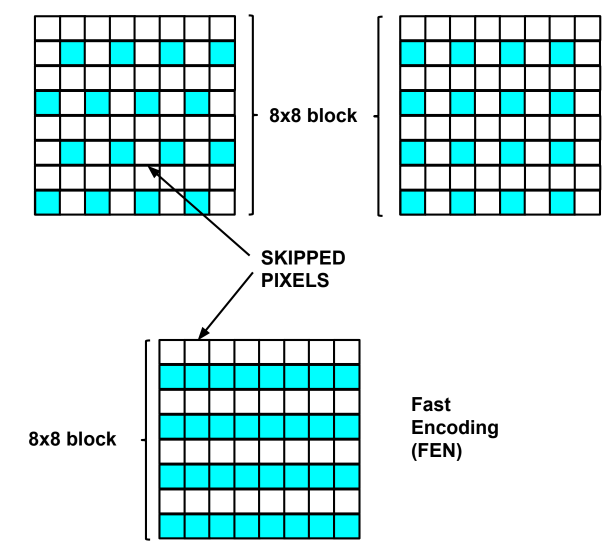


__Figure 2.__ Example of FEN subsampling (bottom) and two further pixel decimation patterns with FEN subsampling.

	B-Slices (Inter)		I-Slices (Intra)		Avg. Time Savings (%)
Sequence	BD Y-PSNR (db)	BD-Rate (%)	BD Y-PSNR (db)	BD-Rate (%)	
PeopleOnStreet_2560x1600_30	0.92	3.2	0.14	3.12	24.9
Kimono1_1920x1080_24	0.04	1.5	0.04	4.21	22.3
Tennis_1920x1080_24	0.09	1.2	0.07	1.18	18.6
BasketballDrill_832x480_50	0.13	2.1	0.11	2.36	25.4
KristenAndSara_1280x720_60	0.05	3.5	0.09	1.24	23.1
FourPeople_1280x720_60	0.14	5.8	0.21	2.22	26.8
BQMall_832x480_60	0.56	3.6	0.07	2.57	19.1

Results A: Intra and Inter mode SSE & SAD Pixel Decimation effects with FEN subsampling enabled.

Figure **??** compares the row-skipping subsampling scheme in FEN with two alternate patterns. Top right is more closely aligned to the work in [**?**], which is a simpler scheme (no shifting of skipped PEL indices), but leaves entire columns as well as rows unaddressed by the cost calculation. The information loss is offset somewhat in the inteleaving pattern top left.

	BD-Y-PSNR (db)	BD-Rate (%)	ATS (%)
Sequence			
PeopleOnStreet	0.01	0.25	16.2
Kimono1	0.01	<0.01	13.3
Tennis	0.02	0.05	15.1
BasketballDrill	0.04	0.3	16.7
KristenAndSara	0.02	< 0.01	11.2
FourPeople	0.02	0.43	17.5
BQMall	0.03	0.41	13.6

Results B: Inter mode SAD Pixel Decimation

Experimentation

We present first the set of conclusive encoding results in Table A. For this set, our pixel decimation modifications to both SSE and SAD distortion calculations are applied, with FEN active, compared to the default optimized case (FEN only). Results are encouraging with time savings averaging 20-25%, bitrate penalties mostly limited to the 3% level, and acceptable PSNR drop. Note this is in addition to the complexity reduction (and associated affects) achieved with the FEN optimization itself. We are, ultimately, testing the applicability of the PD method on top of current tools. And since SSE decimation is included, intra frames are affected, through the cost function J_{mode}.

Next, Table B provides insight into the effects of decimating D_{SAD} alone, with FEN set to 0. We note the heavy contrast between BD-rate penalties here and in Table 1 – changing D_{SAD} does not affect the J_{mode} cost function, thereby the manipulation of D_{SAD} alone, and without the row-skipping functionality of FEN, has comparatively miniscule impact on the birate and quality. The importance and centrality of D_{SSE}-based mode decisions here is highlighted.

Preliminary testing with SAD decimation in a FEN-disabled encoder reveals average time gains in the 12-15% range, implying that the overall combined improvement of using the interleaving pixel decimation patterns presented with the default FEN optimization enabled, is close to 40% of the total encoding time.

We note that neither video resolution nor frame rate were directly correlated with penalties in quality, or with the time complexity improvement. The content of the video streams however, did play a role in determining which sequences benefited the most, although changing scenes in different parts of some sequences did not leave clear indication of this in sequences with highly varied scenery. Relatively static sequences (like FourPeople_1280x720_60) and single-scenery type sequences with movement (BasketballDrill_832x480_50) showed strong time saving gains. We also note that results extended to 2K-resolution video: PeopleOnStreet approaching 25% encoding time reduction in the combined scheme.

Standard JCT-VT recommended test procedures are followed in all the above; test sets conducted at QP $\in \{22, 27, 32, 37\}$ for the calculation of BD-bitrate and BD-PSNR values from the corresponding PSNR/Bitrate curves.

Conclusion

Results presented show the applicability of pixel decimation schemes to SAD and SSE cost functions in the HEVC encoding process. Overall PSNR and bit-rates are minimally affected by the changes to the HM Test Model, with a 4-to-1 subsampling achieved at all block sizes 8×8 and above, when used combined with the FEN scheme of row skipping. Further investigation in large block sizes of width 32 and 64 may provide insight in the possibility of additional, size-dependent subsampling in the cost functions concerned. The SATD, widely used in both modes of encoding, may also provide significant room for improvemnt if appropriate methods are found, particularly for blocks in larger sizes.

References

[1] High Efficiency Video Coding, document ITU-T Rec. H.265 and ISO/IEC 23008-2 (HEVC), ITU-T and ISO/IEC, Apr. 2013.

[2] K. Podder et al., Fast Intermode Selection for HEVC Video Coding Using Phase Correlation, Proc. Intl. Conf. on Digital lmage Comp. Tech and Applic. (DICTA), pg. 1-8. (2014).

[3] Liquan Shen, Zhaoyang Zhang, and Zhi Liu, Adaptive Inter-Mode Decision for HEVC Jointly Utilizing Inter-Level and Spatiotemporal Correlations, IEEE Trans. Circuits Syst., 24, 10 (2014)

[4] JCT-VC, High Efficiency Video Coding Test Model 15 (HM15) Reference Software, JCTVC-Q1002. (2014).

[5] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, Overview of the high efficiency video coding (HEVC) standard, IEEE Trans. Circuits Syst. (Dec. 2012).

[6] Jian-Liang Lin, Yi-Wen Chen, Yu-Wen Huang, and Shaw-Min Lei,

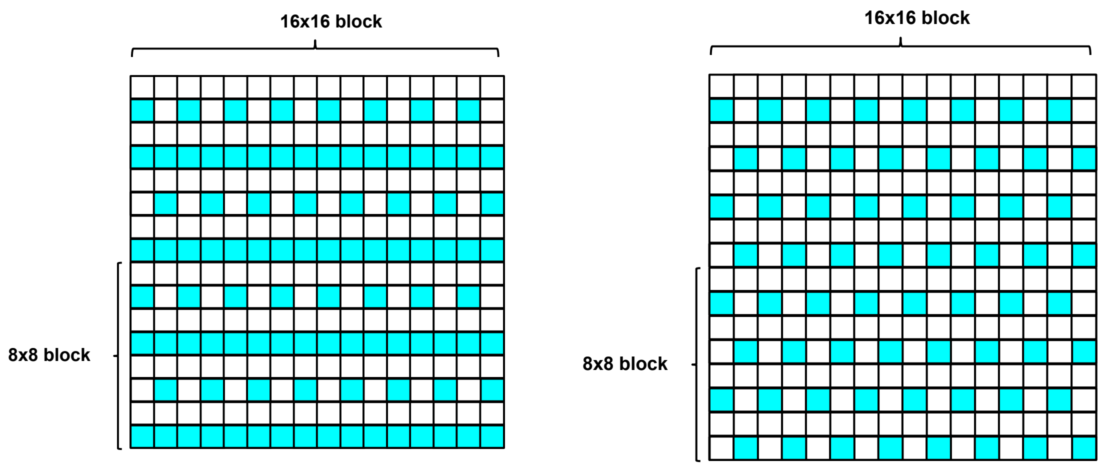
Motion Vector Coding in the HEVC Standard, IEEE J. of Sel. Topics in Sig. Proc., 7, 6, pg. 957-968 (Dec. 2013).
[7] J. Vanne et al., Efficient Mode Decision Schemes for HEVC Inter Prediction, IEEE Trans. Circuits Syst., 24, 9, pg. 1579 (Sep. 2014).
[8] H. Kibeya et al., TZSearch Pattern Search Improvement for HEVC Motion Estimation Modules, Proc. ATSIP, (2014).
[9] B. Liu and A. Zaccarin, New fast algorithms for the estimation of block motion vectors, IEEE Trans. Circuits Syst., 3, 2 (April 1993).

Author Biography

Ahmed Hamza is a PhD student in the School of Computing at the University of Portsmouth. He obtained a M.S in Computer Science from Georgetown University in 2010, where he worked on algorithms in chem-informatics. His interests are in Video Coding and optimization, algorithms, information theory, and natural systems. Ahmed is a member of IET and IEEE.

Abdelrahman Abdelazim is an assistant professor of Computer Engineering at the American University of the Middle East (AUM) in Kuwait. He holds a BEng (Hons) degree in Digital Communication and a PhD degree in Engineering, both from the University of Central Lancashire (UCLAN), Preston, UK. Between 2008 and 2012 he worked as Lecturer in Electronics within the School of Computing, Engineering and Physical Sciences (CEPS). His experience includes leading the implementation of a number of academic and industrial digital communication projects. His research interests are in the area of reducing the complexity of Video Coding Encoders in real-time Scalable and Multi-view applications and the area of teaching and learning in higher education. Abdelrahman is member of the IET since 2006, and he established the Kuwait community in 2013. He is a chartered engineer and a Fellow of the Higher Education Academy.

Djamel Ait-Boudaoud joined the University of Portsmouth in 2010 and is currently Professor and Dean of the Faculty of Technology. Before Portsmouth he was the head of the School of Computing, Engineering and Physical Sciences at the University of Central Lancashire for close to 10 years. His research interests are predominantly focused on the problems of optimisation with applications in 3D computer vision, video standards (H264) and solving combinatorial (ordered-sequence) problems using evolutionary algorithms. He gained a PhD in 1991 from the Department of Electrical and Electronic Engineering at University of Nottingham, UK., is a Chartered Engineer (CEng) and a fellow of the Institution of Engineering and Technology (FIET).



(a) (b)

Figure 3. *Comparison of less aggressive subsampling* **??** *combined with FEN, and our chosen PD scheme* **??**.