ENHANCED INTER-PREDICTION USING MERGE PREDICTION TRANSFORMATION IN THE HEVC CODEC

Saverio G. Blasi ⋆, *Eduardo Peixoto* † *and Ebroul Izquierdo* ⋆

⋆School of Electronic Engineering and Computer Science, Queen Mary University of London
† Departamento de Engenharia Elétrica, Universidade de Brasília

ABSTRACT

Merge prediction is a novel technique introduced in the HEVC standard to improve inter-prediction exploiting redundancy of the motion information. We propose in this paper a new approach to enhance the Merge mode in a typical HEVC encoder using parametric transformations of the Merge prediction candidates. An Enhanced Inter-Prediction module is implemented in HEVC using Merge Prediction Transformation (MPT), integrated with the HEVC new features such as the large coding units (CU) and the recursive prediction unit partitioning. The MPT parameters are quantised according to the CU depth and the current QP. The optimal quantization steps are derived via statistical analysis as illustrated in the paper. Results show consistent improvements over conventional HEVC encoding in terms of rate-distortion performance, with a small impact on the encoding complexity and negligible impact on the decoding complexity.

Index Terms— Inter-prediction, video coding, HEVC

1. INTRODUCTION

The high efficiency video coding (HEVC) is a recently approved video coding standard originated as a response to the Joint Call for Proposals issued by ITU-T and MPEG [1], due to become the successor to the current state-of-the-art H.264/AVC standard [2]. While the architecture of a HEVC encoder remains largely unchanged with respect to its predecessor, it includes many new coding tools, and almost all of the encoder blocks are different and optimised with respect to their counterparts in the H.264/AVC. This allows the new standard to achieve up to 50% bitrate reduction compared to the H.264/AVC [3].

While in H.264/AVC a frame is divided in macroblocks of fixed size (equal to 16×16 pixels), the HEVC encoder makes use of a flexible partitioning where frames are divided in a recursive way in entities called coding units (CU). The encoder always consider the largest CU (LCU) of fixed size (usually equal to 64×64 pixels), which is assigned a depth equal to 0. This is then divided in four smaller CUs at depth 1, and so on recursively up to a minimum CU size of 8×8 pixels. The inter-prediction module is still based on multiple reference frames and variable block-size Motion Estimation (ME). In particular when testing inter-prediction on a given CU (at a certain depth), this is further divided in up to four smaller entities called prediction units (PU), according to a set of possible modes. A typical encoder tests all possible sub-partitioning of an LCU, performing motion estimation (ME) for each mode and finally selects the best configuration for the LCU in terms of coding efficiency.

Email: sgblasi@eecs.qmul.ac.uk, eduardopeixoto@ieee.org, ebroul.izquierdo@eecs.qmul.ac.uk.

This research was partially supported by the European Commission under contract FP7-247688 3DLife and FP7-287723 REVERIE

We previously [4] introduced the concept of the enhanced inter-predictor (EIP) with shifting transformation (ST) in the context of ME in H.264/AVC, proving that using a local transformation in the inter-prediction candidates can successfully increase the coding efficiency. In this paper we instead investigate the effects of local transformations on the Merge estimation of the HEVC. In particular, a Merge prediction transformation module (MPT) is implemented and used along conventional Merge estimation to transform the Merge candidates reducing the prediction error. Due to the characteristics of the HEVC codec, such as the large PU sizes and especially the internal bit depth increase (IBDI) [5], an adaptive quantization step (QS) needs to be included in the MPT to reduce the set of possible parameter values hence reducing the number of required bits for encoding the parameter. The QS is found as a function of the CU depth and current QP following from a statistical analysis on test data. Results show that the MPT can efficiently enhance the coding efficiency of a conventional HEVC encoder while having a very modest impact on both encoding and decoding complexity.

2. RELATED WORK

The HEVC includes several tools which aim at improving coding efficiency using pixel transformation. Sub-pixel ME [6] based on reference frame interpolation and fractional motion vectors (MV) up to quarter precision is implemented in the standard similarly to the H.264/AVC. Despite the larger amount of bits required to transmit sub-pel MV components, more efficient compression can be achieved due to the smaller resulting residuals. Weighted Prediction (WP) is also implemented, with no significant differences from H.264/AVC [3]. When using WP, the reference frames used for ME are transformed using a weighting factor and additive offset. Two WP modes are allowed [7], namely explicit and implicit modes. In the former, the parameters are computed for each reference and then transmitted in the slice header, whereas in the latter the parameters are not transmitted but extracted as a function of global features of the current macroblock and reference frames. Finally, another tool is introduced in HEVC at the end of the encoding loop that also makes use of frame transformation, called Sample Adaptive Offset (SAO) [8]. The SAO is an additional filter aimed at reducing the distortion between the current original and reconstructed frames, by means of an additional offset applied to the reconstruction after de-quantization and de-blocking filter. The SAO parameters are calculated using information available after the current frame is encoded, and are then encoded in the bitstream. At the decoder side, the relevant offset is applied to all the pixels in a frame in the same category.

The Merge mode is introduced in HEVC [9] as a novel technique to increase the efficiency of the inter-prediction. It consists of selecting as the motion information (MI, comprising the motion vectors (MV), prediction direction and possibly one or two reference indexes) for the current PU, the MI of a previously encoded PU (i.e.,

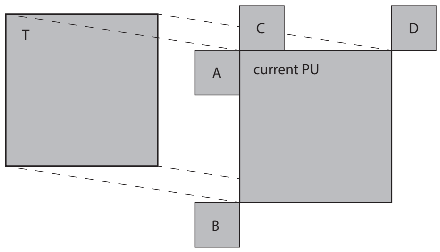

Fig. 1. Merge prediction candidates.

"merging" this MI). While no MI needs to be encoded when using this technique (apart from an index to extract the correct previously encoded PU), the encoder might still get sufficiently accurate predictions due to the presence of spatial and temporal redundancy in the sequence. Experiments show [9] that an average 8% (and up to 20%) BD-rate reductions [10] are obtained using this technique. Merge prediction shares some similarities with the SKIP mode (in fact both methods rely on a prediction of the MI from previously encoded units), but it is worth noticing that while SKIP is only considered on full CUs and no residual coding is performed, the Merge mode is allowed on all PUs and the resulting residuals are conventionally encoded and transmitted.

In particular, before performing inter-prediction on the current PU, a list of up to five Merge candidates is populated with the MI of previously encoded PUs (as shown in Figure 1). Four neighbouring PUs are considered in the current frame, located at positions A, B, C and D in the figure. Additionally, the PU in a previously encoded frame at the same location as the current PU is considered, referred to as T in the figure (the first frame considered as reference for ME is used for this purpose). If any of these candidates are not available, they are not included in the Merge candidate list.

While encoding a PU, conventional ME is performed to find the optimal MI in terms of a rate-distortion (RD) cost, which takes into account both the prediction distortion and the cost due to the rate (in bits) required to transmit such MI. This RD cost is compared with the cost of the merge candidates. The cost of each candidate is computed using the same RD measure, comprising the distortion obtained using the candidate MI, plus the cost of transmission of the index needed to select that candidate. Notice that this cost is extremely small compared to the cost of transmission of the ME solution, because a single unsigned integer smaller than 4 needs to be transmitted instead of the entire MI. Finally if the cost of any Merge candidate is lower than the ME cost, the candidate with minimum cost is chosen and used for inter-prediction for the current PU. A flag is transmitted per PU to signal whether the Merge mode is used.

Merge prediction results in a very efficient encoding, but unfortunately it is only selected on those PUs where suitable candidates can be found that provide sufficiently accurate predictions: in average the Merge mode represents less than 15% of the total prediction modes selected in conventional HEVC [9]. Previously [4] we showed that local transformations can be successfully applied on certain ME candidates to enhance the inter-prediction accuracy improving the coding efficiency, in the scope of a H.264/AVC codec. On the light of these results, a similar approach is presented here where local transformations are applied on the Merge candidates to reduce the distortion, and thus enhancing prediction accuracy. In order to do so the HEVC encoder was modified to include the proposed MPT module, as it is shown in the following section.

3. MERGE PREDICTION TRANSFORMATION

In conventional HEVC the distortion of each Merge candidate is computed making use of the same error metric used for ME (usually the SAD) between the pixels in the PU and the pixels in the inter-prediction obtained using such Merge candidate. In particular denote the current PU as $\mathbf{T}$, and consider a Merge candidate k and the associated inter-prediction $\mathbf{P_k}$ resulting from using such candidate. Refer to each element in $\mathbf{T}$ as $T(i)$, for $i = 0, 1, ..., (B-1)$ where $B = M \times N$, and M and N are the PU height and width respectively. Similarly $P_k(i)$ are the elements in $\mathbf{P_k}$. If we denote the difference vector between $\mathbf{T}$ and $\mathbf{P_k}$ as $\mathbf{D_k}$, we can define the SAD of the current Merge candidate as:

$$SAD\left(\mathbf{T}, \mathbf{P_k}\right) = \sum_{i=0}^{B-1} \left|D_k\left(i\right)\right| = \sum_{i=0}^{B-1} \left|P_k\left(i\right) - T\left(i\right)\right| \quad (1)$$

In order to transform the prediction candidates, we make use of a simple transformation function consisting of a single parameter uniformly added to all pixels in a prediction block. Formally, denote as $\hat{\mathbf{P}}_\mathbf{k}$ the prediction obtained using MPT on $\mathbf{P_k}$, where:

$$\hat{P}_k(i) = P_k(i) + s, i = 0, 1, ..., (B-1) \quad (2)$$

We are interested in finding the optimal parameter s such that the enhanced SAD:

$$SAD\left(\mathbf{T}, \hat{\mathbf{P}}_\mathbf{k}\right) = \sum_{i=0}^{B-1} \left|P_k\left(i\right) + s - T\left(i\right)\right| \quad (3)$$

between the original PU and the MPT prediction as in Eq. 2 is minimum. Consider now with no loss of generality that the elements in $\mathbf{D_k}$ are rearranged in increasing order, i.e. $D_k(i) \leq D_k(j), \forall i \leq j$. Denote also with N_+, N_0 and N_- the number of positive, zero and negative elements in $\mathbf{D_k}$ respectively, which we refer collectively as the sign ratio. We proved [4] that the optimal parameter s for a given $\mathbf{P_k}$ has a closed form solution. In particular, refer to the following as Condition 1:

$$N_-\left(0\right) > N_0\left(0\right) + N_+\left(0\right) \quad (4)$$

If Condition 1 is satisfied for the current couple $\mathbf{T}$ and $\mathbf{P_k}$, then the optimal MPT parameter is a negative integer, which can be found as:

$$s = -D_k\left(N_- - n_{max}\right) \quad (5)$$

where n_{max} is defined as:

$$n_{max} = \left\lfloor \frac{N_- - N_0 - N_+}{2} \right\rfloor \quad (6)$$

A solution can also be found under a similar condition resulting in a positive optimal MPT parameter. Note that this covers all possibilities, since the optimal parameter s is either positive, negative or zero. Finally the enhanced SAD as a result of the transformed prediction can also be computed in a single step using a closed-form solution. Using the above formulas, we can find the optimal MPT parameter for each Merge candidate with its associated enhanced distortion.

In order for a PU predicted using the Merge mode to benefit from the MPT, the optimal parameter s needs to be transmitted in the bit-stream along with the information required to correctly decode the motion information (i.e. the Merge index). Notice that, while a PU encoded with Merge can be theoretically used as a Merge candidate for other PUs, the MPT parameters instead are never merged and are computed and transmitted independently for each Merge candidate. For this reason, the RD cost of a certain Merge candidate k is modified to include such necessary additional information as follows:

$$J_k = SAD\left(\mathbf{T}, \hat{\mathbf{P}}_\mathbf{k}\right) + \lambda_{MRG} R_k + \lambda_S R_s \qquad (7)$$

where λ_{MRG} and λ_S are the Lagrangian multipliers used for computing the RD cost of the Merge index and the MPT parameter respectively.

Following the conventional HEVC coding loop, the encoder analyses CUs at different depths, testing for each CU several inter-prediction modes on PUs of different sizes. Each time the Merge mode is tested, the MPT module is used to enhance the prediction returning the transformed prediction and the optimal MPT parameter for the current PU. If the Merge mode is chosen after mode decision, such parameter is finally encoded in the bit-stream.

The modified HEVC with MPT was initially tested on some video sequences from the HEVC test sequences, to analyse the distribution of the resulting MPT parameters. According to these results, the optimal MPT parameters follow a distribution centered in zero where high values are less probable. For this reason they are encoded in the proposed modified HEVC encoder using variable-length coding (VLC) tables [11], which require few bits for small values and progressively more bits for higher, less probable values. Also for the same reason, λ_S in Eq. 7 is set equal to the Lagrangian multiplier used for computing the RD cost for the MV components.

Tests have shown that while high gains in terms of prediction accuracy can be obtained using the transformation, the encoder efficiency is limited by the high number of bits needed to encode the MPT parameters. This is particularly important in the case of the HEVC if the internal bit depth increase (IBDI) technique is being used [5]. In this case, the encoder (and the decoder) artificially increases the word-length of every pixel by two bits (by default), and performs all internal computations using this more accurate representation. Note that this highly enlarges the range of possible MPT parameters and, since these are being encoded using a VLC table, severely impacts the coding efficiency of the MPT technique.

In order to address this problem, the MPT was modified by reducing the number of possible parameter values consequently decreasing the amount of bits required for the encoding. In particular, instead of allowing the MPT parameter to assume all the integer numbers, a certain quantization step (QS) is considered, such that parameters are forced to only assume multiple values of the QS. The algorithm in this section was modified to include this modification returning the best MPT parameter for a PU using a given QS. Notice that, being the SAD as a function of the MPT parameter a convex function, the optimal quantized parameter in terms of the SAD is such value at the given QS that is nearest to the optimal non-quantized parameter. For this reason we can assume a similar behaviour when including the RD cost in the selection of the optimal MPT parameter (as in Eq. 7). Comparably low costs can be expected from the enhanced Merge prediction using quantized MPT with appropriate QS.

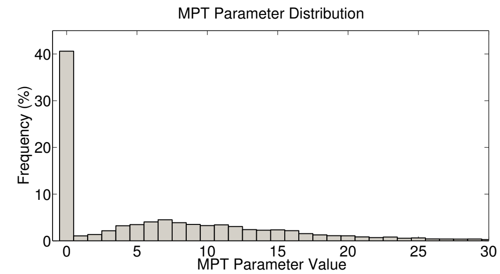

Fig. 2. Histogram of the MPT parameter values.

From these initial results we also observed that the optimal MPT parameters strongly depend on local features of the current PU being encoded, particularly the PU size and the encoding QP. For these reasons, it makes sense to locally adapt the QS using different values depending on such features. A statistical analysis of the correlation between optimal MPT parameters and local features of the PUs was used based on experimental tests on actual video sequences in order to derive the best QS value for a given PU.

4. OPTIMAL QS MATRIX DERIVATION

Tests were performed on the first 50 frames of three HEVC test sequences at WVGA (832×480) resolution (namely *PartyScene*, *RaceHorses* and *Keiba*). To reduce the complexity of the analysis, the CU depth was used as the classifier for the PU size (with only four possible values from 0 to 3). In each test the MPT was enabled only in PUs extracted from CUs at a given depth, while conventional Merge mode was used in all other cases. As long as the QP is concerned, four values were used (namely 22, 27, 32 and 37). All 16 depth/QP combinations were tested.

Results immediately show that there is a strong correlation between the depth/QP combination and the number of zero-valued MPT parameters. The percentage of zero-valued parameters over the total number of PUs encoded with Merge mode is shown in Table 1. Clearly smaller PUs (i.e. extracted from CUs at depth 3) and higher QPs resulted in a very high percentage of zero-valued parameters (as high as 74% in the case QP$=$ 37). Notice that a zero-valued parameter corresponds to the case when the MPT is ineffective on a certain PU (no transformation could be found to enhance the Merge prediction). Enabling the MPT in these PUs requires additional bits to be encoded in the bitstream, to signal the zero values, even if no enhancements actually result from using the transformation. Thus, the MPT is disabled in PUs corresponding to a percentage of zero-valued parameters greater than 40%.

Table 1. Percentage of zero-valued MPT parameters.

		Depth			
		0	1	2	3
QP	22	30.15	28.45	35.26	57.86
	27	24.08	29.07	32.85	66.61
	32	22.63	32.99	50.79	72.07
	37	26.49	40.60	57.71	74.12

The plot in Figure 2 shows the histogram of the absolute values of optimal MPT parameters in the case of depth 0, $QP = 37$. Notice that quantising the MPT parameters would result in little loss of prediction accuracy but a high increase in coding efficiency: for instance, while 7 bits are needed to encode the signed integer 4 using $QS = 1$, only 3 bits are needed if $QS = 4$. It should also be considered that larger parameters usually yield a larger impact on

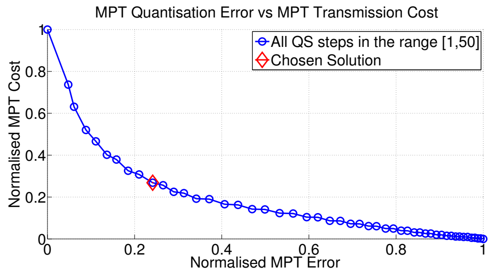

Fig. 3. MPT Quantisation Error vs MPT Transmission Cost.

the SAD and, therefore, quantizing the parameters allows large values to be chosen more often (as these values would now use smaller codewords). Following from this example it is clear that a more efficient encoding can be achieved quantising the MPT parameters as long as the additional distortion due to the quantization would remain below a certain limit. Hence, in order to find the best QS for a given depth/QP combination, we analysed how the MPT parameter cost and the MPT quantization error relate to the QS.

In the following we refer to non-quantized MPT parameters as s_1 parameters, and correspondingly to parameters found using a certain $QS = X$ as s_X parameters. Values of QS between 1 and 50 were considered in the analysis. The approximation error resulting from using a quantization $X \neq 1$ was estimated assuming that the s_1 parameters are always approximated to the nearest multiple value of X. Correspondingly we assumed that no error happens at $QS = X$ if a s_1 parameter is a multiple value of X. We also assumed the error to be dependent on the distance between the s_1 parameter and its s_X approximation, thus we computed the total estimated error as the number of s_1 parameters in the test data multiplied by their absolute difference with the nearest multiple value of X. On the other hand, the number of bits required to encode each quantized MPT parameter was computed for each considered value of the QS. This was then normalized to the total number of parameters to obtain the average word-length per PU to encode the MPT parameters at a given quantization.

Figure 3 shows the results obtained with depth equal to 0, $QP = 37$. For each value of the QS, the estimated error was plotted against the average number of bits. Finally, we selected as the optimal QS for this depth/QP combination the value that corresponds to the point in the plot which is nearest to the origin in terms of Euclidean distance (6 in the example). The process was repeated for each depth/QP combination to obtain a matrix of optimal QS values, which is shown in Table 2.

Table 2. The optimal QS matrix. Note that the MPT is disabled in some depth/QP combinations.

		Depth			
		0	1	2	3
QP	22	4	4	6	N/A
	27	4	6	8	N/A
	32	4	8	N/A	N/A
	37	6	N/A	N/A	N/A

Finally the encoder was modified to make use of the QS matrix. Four ranges are considered for the QP, centered respectively in 22, 27, 32 and 37. The encoder classifies a PU according to such ranges and the depth of the CU and extracts the corresponding QS. If Merge prediction is used on the PU, quantized MPT is then performed using this quantization.

5. RESULTS

The approach was tested on several sequences from the HEVC test set using the low-delay B configuration. Full results are presented in Table 3 in terms of the BD bitrate [10] (i.e., the average bit-rate difference relative to an anchor, in percentage) , where the HM reference software (release candidate 6.0 [12]) is used as anchor for all tests. The sequences used can be seen in Table 3.

The modified encoder with MPT outperforms conventional HEVC in all cases, obtaining an average BD bitrate reduction of 0.6% and up to 2.6% in one particular sequence. This peak in the MPT performances, obtained for the Mobisode2 sequence, is due to the particular features of such sequence which presents largely static content but drastic brightness variations, allowing the Merge mode to be chosen frequently and the MPT to be very effective at the same time. In all other cases reductions around 0.3% are obtained, generally distributed towards considerable reductions in the bitrate and unaffected reconstruction PSNRs, reflecting what was obtained with the EIP in H.264/AVC.

The tests validate the approach showing that even considering the extremely limited scope of the MPT, still the transformation has a clear impact on the encoder performances. Notice in fact that Merge estimation is in average used in less than 15% of the prediction modes while encoding in conventional HEVC [9]. Notice also that the high complexity of the HEVC codec provides extremely efficient coding performances, and therefore modifications to the codec usually result in smaller gains than those obtained in less efficient standards. For these same reasons though the MPT has a very small impact on the coding complexity: the modified encoder requires only around 5% additional time to complete the encoding, while the MPT has negligible effects on the decoding time (in average less than 1% increase).

Table 3. BD rate of the MPT approach against conventional HEVC.

Resolution	Sequence	FPS	BD-rate MPT
1024×768	ChinaSpeed	30	-0.5
832×480	BQMall	30	-0.2
	Keiba	30	-0.1
	PartyScene	50	-0.3
	RaceHorses	30	-0.2
416×240	BlowingBubbles	50	-0.4
	Mobisode2	30	-2.6

6. CONCLUSIONS

An enhancement to the Merge prediction in conventional HEVC is presented in this paper making use of pixel transformation. In particular, Merge prediction transformation is performed on the Merge candidates to increase the prediction accuracy at the cost of transmission of the transformation parameters. Consistent gains are obtained in all tests with average BD-bitrate reductions of 0.6% and up to 2.6%, even considering the relatively small impact of the technique in terms of the percentage of affected PUs.

In the light of these results, in the future more complex transformation methods can be investigated to further enhance the prediction accuracy. The approach can also be extended to other prediction modes subject to an appropriate analysis of the highly optimised HEVC features.

7. REFERENCES

[1] "Joint Call for Proposals on Video Compression Technology", document VCEG-AM91 of ITU-T Q6/16 and N1113 of JCT1/SC29/WG11, Kyoto, Japan, 2010

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*, v. 13, n.7, pp. 560-576, 2003.

[3] G. J. Sullivan, J. R. Ohm, W. J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard", *IEEE Transactions on Circuits and Systems for Video Technology*, 2012.

[4] S. G. Blasi, E. Peixoto and E. Izquierdo, "Enhanced Inter-Prediction via Shifting Transformation in the H.264/AVC", IEEE Transactions on Circuits and Systems for Video Technology, 2012.

[5] T. Chujoh and R. Noda,"Internal bit depth increase for coding efficiency", ITU-T SG16/Q6, VCEG Meeting, Doc. VCEG-AE13, 2007.

[6] B. Girod, "Motion-compensating prediction with fractional-pel accuracy", *IEEE Transactions on Communications*, v. 41, n. 4, pp. 604-612, 1993.

[7] J. M. Boyce, "Weighted prediction in the H.264/MPEG AVC video coding standard," *International Symposium on Circuits and Systems*, v.3, pp. iii-789-92, 2004.

[8] C.-M. Fu, C.-Y. Chen, Y.-W. Huang and S. Lei, "Sample adaptive offset for HEVC", *IEEE 13th International Workshop on Multimedia Signal Processing*, pp.1-5, 2011.

[9] P. Helle, S. Oudin and B. Bross, "Block Merging for Quadtree-Based Partitioning in HEVC", *IEEE Transactions on Circuits and Systems for Video Technology*, 2012.

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

[11] I. Richardson, "White Paper: H.264/AVC Context Adaptive Variable Length Coding", Vcodex , 2011.

[12] JCT-VC, "Subversion Repository for the HEVC Test Model version HM6.0rc", https://hevc.hhi.frauhnofer.de/svn/svn_HEVCSoftware/tags