SMART DECODER: A NEW PARADIGM FOR VIDEO CODING

D-K. Vo-Nguyen[1,2], J. Jung[1], J-M. Thiesse[3], M. Antonini[2]

[1]Orange Labs, [2]I3S Laboratory (University of Nice Sophia-Antipolis - CNRS), [3]ATEME

ABSTRACT

The coding efficiency of the new video coding standard, High Efficiency Video Coding (HEVC), is strongly associated with better use of spatio-temporal redundancies thanks to an increased number of competing coding modes. However, this competition involves a massive increase in signaling bitrate which becomes a possible limit for the next generation of encoder.

This paper proposes a new coding scheme that breaks with conventional approaches. It exploits a more complex decoder able to reproduce the choice of the encoder based on causal references, eliminating thus the need to signal coding modes and associated parameters. The general outline of this new codec and a proposed implementation are described in this paper. Experimental results under common test conditions report an average bitrate saving of 1.7% at the same quality compared to HEVC for a wide range of video sequences.

Index Terms— HEVC, Smart Decoder, Video Coding

1. INTRODUCTION

High Efficiency Video Coding (HEVC) [1], successor of H.264/MPEG-4 AVC, is the new video coding standard developed jointly by ISO/IEC MPEG and ITU-T VCEG through the "Joint Collaborative Team on Video Coding" (JCT-VC) group. As its predecessors, the compression ability of HEVC comes from the use of Inter and Intra predictions that exploit redundancies respectively in time and space. Compared with previous standards, HEVC features an increased number of competing coding modes and their associated parameters [2]. For instance, quad-tree based partitioning allows to split a coding tree unit (CTU) in different coding units ranging from 64×64 down to 8×8 pixels; Intra prediction includes up to 35 directions against only 9 in H.264/AVC; for Inter prediction, AMVP (Advanced Motion Vector Prediction) and Merge modes have respectively two and five predictors in competition. The introduction of these new coding modes and parameters allows a more accurate prediction but generates a significant increase in signaling bitrate. In the future, more coding modes could be added and generate an overhead that could limit the compression performance.

To overcome this limit and further improve the coding efficiency, several approaches have been proposed and can be classified into two main categories. The first one aims to reduce the signaling cost of encoding information, especially with the use of "Most Probable Mode". In [3] and [4], it is proposed to dynamically reduce the number of Intra predictors. In [5], optimal Intra predictor is adaptively determined based on causal pixels only, thus saving the associated signaling. The second category includes approaches that exploit the complexity of the decoder to derive the choice of the encoder, thereby eliminating the need for signaling. We classify them as "Smart Decoder" (SDec) based approaches, since the decoder has the ability to compute decisions made by the encoder. Some of them are based especially on "Template Matching" process that exploits the causal surrounding area of a block to derive motion vectors without signaling. This principle applied in Inter [6, 7, 8] proposes motion estimation that is jointly performed in both the encoder and the decoder in order to reduce motion vectors' transmission cost. Other variants can be applied in Intra [9] where the current block takes as predictor a block located in the reconstructed area of the current frame by evaluating the similarity of their causal surrounding pixels. Nevertheless, these methods allow to reduce motion information, not the signaling of coding modes as proposed in this paper.

This paper concerns next codec generations after HEVC in which the number of coding modes in competition is expected to be further increased. To prepare for this perspective where the reduction of signaling overhead becomes crucial, we propose a new coding scheme which aims to reduce the signaling of coding modes and their associated parameters. Similar to the encoder, the decoder is provided the ability to perform the computation of optimal coding modes based on a causal reference, removing therefore the need to signal them in the bitstream. This new approach can thus be also classified in "Smart Decoder" category.

The rest of this paper is organized as follows: Section 2 describes the general encoding and decoding processes of our SDec scheme, along with its advantages over conventional codecs and a proposed implementation. Section 3 presents experimental results. We finally conclude in Section 4 by specifying different possible perspectives.

2. PROPOSED CODEC DESIGN

2.1. General description of the SDec encoding scheme

In HEVC, a prediction unit (PU) is the basic block for prediction. In conventional encoding scheme, all available coding modes with their associated parameters are set to compete on

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

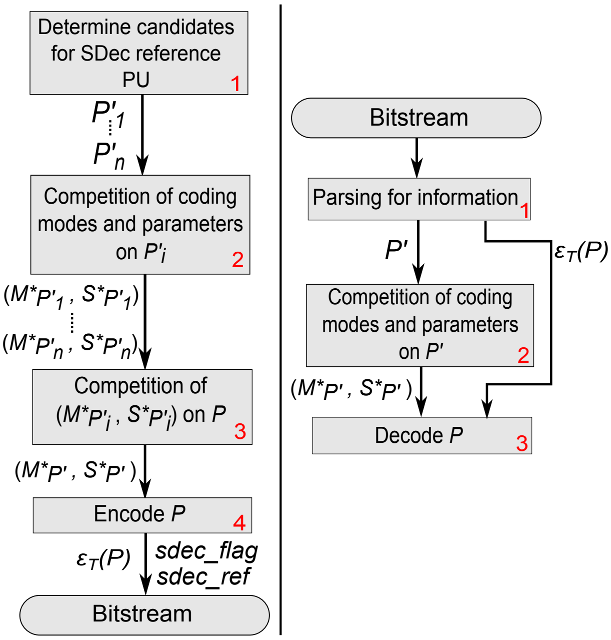

Fig. 1. General SDec encoding (left) and decoding (right) schemes

the current PU. The optimal mode which minimizes the rate-distortion (R-D) cost is then selected to encode the current PU and is signaled in the bitstream. In the proposed encoding scheme, the coding mode selected to encode the current PU is not calculated directly on the latter, but on a causal PU called *SDec reference PU* instead. The similar process of computing the competition between coding modes can be performed by the decoder, avoiding to signal the selected optimal mode in the bitstream.

In the rest of the paper, we use the following notations:

- P the current PU to encode,
- P' the SDec reference PU selected among n candidates $(P'_1, ..., P'_n)$,
- $(M^1, ..., M^m)$ m coding modes, each having associated parameters $(S^{1,M_i}, ..., S^{q_i,M_i})$, (for example, M^1 could be HEVC Intra mode, with $(S^{1,M_1}, ..., S^{q_1,M_1})$ corresponding to different Intra directions: horizontal, vertical, DC...)
- $M^*_{P'}$ and $S^*_{P'}$ the optimal coding mode and parameter selected by SDec for encoding P',
- $Pred_{M^*_{P'},S^*_{P'}}(P)$ the prediction of P using $M^*_{P'}$ and $S^*_{P'}$,
- $\varepsilon_T(P)$ the texture residual after encoding P.

Regarding the information signaled in the bitstream, SDec scheme introduces a syntax element $sdec\_flag$ signaling usage of SDec mode among other available coding modes (Intra, Inter, Merge ...). Another syntax element, $sdec\_ref$ is signaled to indicate which P' is selected. An important characteristic of the proposed scheme is that the coding mode and its associated parameter used to encode the current PU are not transmitted.

The left side of Figure 1 presents the SDec encoding process that can be described in four steps:

Step 1:
In the general SDec scheme, three approaches are possible for selecting P':

- P' is chosen from a list of pre-identified PU candidates. $sdec\_ref$ is then the index of the selected candidate and is transmitted.
- P' is dynamically calculated using a motion estimation. $sdec\_ref$ is then the motion vector pointing to P' and is transmitted.
- P' is a pre-identified PU (such as the colocated block of P). There is no need to transmit $sdec\_ref$.

In a specific implementation of the SDec scheme, one of these approaches, known by the decoder, is used for all PUs.

Note that in the proposed scheme, the selection of P' is crucial because it determines the coding mode and parameter used to encode P. A good reference P' must be correlated with P, so that the coding information calculated on P' is adapted to P and thus provides a good prediction of the latter.

Step 2:
This step computes the optimal coding mode and parameter for each SDec reference PU P'_i. All coding modes and parameters are competing to encode P'_i. More specifically, each pair $(M^j, S^{l,M_j})_{j\in[1..m],l\in[1..q_j]}$ of coding mode and parameter is used to encode P'_i in order to evaluate the optimal pair in terms of R-D cost:

$$J = D + \lambda R$$

where J is the R-D cost, D is the distortion, R is the estimated encoding rate, and λ is the Lagrange multiplier that depends on the quantization parameter. Since coding modes and associated parameters are not transmitted, R does not include their signaling cost. This R-D computation also requires the texture residual resulting from each encoding of P'_i using (M^j, S^{l,M_j}) and is calculated as follows:

$$\varepsilon_T(P'_i) = P'_i - Pred_{M^j,S^{l,M_j}}(P'_i)$$

At the end of this step, all n candidate P'_i have a pair of optimal coding mode and parameter, noted $(M^*_{P'_i}, S^*_{P'_i})$, which minimizes J.

Step 3:
This step determines, among n optimal coding modes of n candidates, the optimal coding mode for the current PU P using R-D based competition. Each optimal pair of coding mode and associated parameter $(M^*_{P'_i}, S^*_{P'_i})_{i\in[1..n]}$ is used to encode P and the optimal pair $(M^*_{P'}, S^*_{P'})$ is selected, with P' the associated SDec reference PU. Note that this step is conducted separately from the step 2 to assure the decodability.

This step is skipped if there is only one candidate for P'.

Step 4:
This step encodes P using the optimal coding mode and parameter $(M^*_{P'}, S^*_{P'})$. A texture residual resulting from the encoding of P is finally calculated to indicate the prediction er-

ror compared to the original PU:

$$\varepsilon_T(P) = P - Pred_{M^*_{P'},S^*_{P'}}(P)$$

2.2. General description of the SDec decoding scheme

At the decoder, $sdec\_flag$ is parsed to check if a PU P is encoded with SDec or not. If it is the case, three steps shown in the right part of Figure 1 are applied:

In the first step, the bitstream is parsed to retrieve data necessary for the SDec decoding process, for example the $sdec\_ref$ flag needed to determine P'.

The second step performs the competition of all coding modes and associated parameters on P'. Simulating the encoder, the decoder encodes P' to find the pair $(M^*_{P'}, S^*_{P'})$ that minimizes the R-D criterion.

In the last step, the current PU is decoded using the optimal coding mode and parameter $(M^*_{P'}, S^*_{P'})$. The texture residual $\varepsilon_t(P)$ extracted from the bitstream is finally added to reconstruct P.

2.3. Advantages and drawbacks of the SDec design

The ability to have an unlimited number of coding modes and parameters in competition without suffering from excessive signaling cost is a major advantage of the SDec scheme. Very different coding modes with their associated parameters can be set in competition, and consequently the adaptation to specific content or characteristic of the scene can be automatically handled.

SDec scheme consequently allows to integrate coding techniques that usually fail due to excessive burden resulting from the large number of their intrinsic parameters in competition, such as "compressed sensing" [10] which requires to transmit transforms and sampling factors, "1D short distance Intra prediction" [11, 12] or "geometry partitioning" [13]. Moreover, complex processes like machine learning become applicable.

The major drawback of the proposed scheme is the additional complexity at the encoder and decoder given that all coding modes and parameters must be tested to find the optimal solution. However, this complexity is fully scalable, in the sense that processing power required to perform a task is allocated according to its need by the codec device. It can be negotiated and adapted on the fly in case of interactive applications or managed by profiles definition.

2.4. Proposed implementation of the general SDec scheme

We propose in this section a specific and simple implementation of the SDec scheme that conforms to the general specification proposed in 2.1 and makes use of some advantages of the SDec design. There are restrictions for selecting the SDec reference PU, on the number of coding modes in competition and on the signaling method for syntax elements.

2.4.1. Selection of candidates for the SDec reference PU

In this paper, we compare two simple methods for selecting candidates for the SDec reference PU P':

<u>SDec 1</u>: using only one pre-identified causal PU candidate as P' ($sdec\_ref$ not signaled, $n = 1$)

We propose to use the colocated PU of P, located in the previous reconstructed image, as the SDec reference PU.

<u>SDec 2</u>: P' is selected among two pre-identified causal PUs candidates ($sdec\_ref$ signaled as index of selected candidate)

Using more candidates increases the likelihood that P' is relevant for coding P; however, it requires bits for signaling the selected candidate. After preliminary tests, we found that having two candidates is a good compromise ($n = 2$). We propose to use the colocated PU of P, and the PU pointed by the motion vector which has its X and Y components as median values among all AMVP motion vector candidates. Both PUs are located in the previous reconstructed frame.

2.4.2. Selection of coding modes for SDec competition

The number of available coding modes tested during the second step of SDec is limited to HEVC Intra mode ($m = 1$). Parameters for Intra mode, which consist of 35 Intra directions, compete thus on each PU candidate ($q = 35$).

2.4.3. SDec mode signaling method

SDec mode competes with existing coding modes and is signaled by the syntax element $sdec\_flag$ for each PU. In this paper, this syntax is signaled by CABAC using three contexts.

3. EXPERIMENTAL RESULTS

3.1. Experimental environment and settings

Our experiments are performed using HEVC software test model version 10.1 (HM10.1). Two versions, with different methods of selecting the SDec reference PU, are tested under JCT-VC common test conditions with the Low delay P profile.

To evaluate the performance of proposed SDec application, sequences from the HEVC standard test set are used. They are classified according to the resolution (Class A 2560×1600, class B 1920×1080...) or the visual content (class F). Additional sequences known for their challenging content are also tested. By varying the quantization parameter (QP) value, three ranges of QP are selected to study SDec performance under different coding rates: low bitrate (LBR): QP 27-32-37-42, medium bitrate (MBR) : QP 22-27-32-37 and high bitrate (HBR): QP 17-22-27-32. This yields to three quality levels with average peak-signal-to-noise ratio for luminance of 33.5 dB, 36.4 dB and 39.5 dB respectively, corresponding to bitrates relevant for some applications. Note that gains are measured using Bjøntegaard Delta rate [14] which represents the average difference between two R-D curves on the considered QP range.

3.2. Coding gains

Table 1 gives the gain in luminance of the two tested versions under three considered QP ranges. Both versions provide systematic gain for all tested sequences.

On the JCT-VC test set, we observe average gains of 1.2%, 0.7% and 0.5% respectively in LBR, MBR and HBR when comparing *SDec 1* to the default HM10.1. Gains up

Sequences class	*SDec 1*			*SDec 2*		
	LBR	**MBR**	**HBR**	**LBR**	**MBR**	**HBR**
Class A	1.7	1.1	0.6	2.0	1.3	0.7
Class B	1.2	0.7	0.4	1.3	0.8	0.5
Class C	0.9	0.5	0.3	1.2	0.8	0.4
Class D	0.8	0.4	0.2	1.0	0.5	0.3
Class E	1.2	0.8	0.6	1.2	0.8	0.7
Class F	1.3	1.0	0.7	1.3	1.0	0.7
Average	**1.2**	**0.7**	**0.5**	**1.3**	**0.9**	**0.5**
Max gain	2.8	2.3	1.8	3.3	2.2	1.7
RollingTomatoes 1088p	2.2	2.1	1.6	2.2	2.5	1.8
PedestrianArea 1088p	2.0	1.4	1.0	1.9	1.5	1.1
Movie 1088p	1.3	1.0	0.8	1.3	1.3	1.0
RushHour 1080p	2.8	1.9	2.1	2.8	1.9	1.5
Tennis 1080p	1.9	1.1	0.7	2.0	1.3	0.8
Crew 720p	2.2	1.4	0.8	2.4	1.7	1.0
Average	**2.1**	**1.5**	**1.2**	**2.1**	**1.7**	**1.2**

Table 1. Bitrate savings in percentage of *SDec 1* and *SDec 2* compared to default HM10.1

to 2.8%, 2.3% and 1.8% respectively for those three encoding rates are achieved. On additional sequences, *SDec 1* also performs very well, with 2.1%, 1.5% and 1.2% of gain respectively in LBR, MBR and HBR. We observe that SDec is very efficient in LBR and that it suits particularly for sequences containing film grain or complex motion.

With the second version *SDec 2* where SDec reference PU is chosen among two candidates, a slightly higher average gain is obtained: 1.3%, 0.9%, 0.5% for JCT-VC test set and 2.1%, 1.7%, 1.2% for set of additional sequences in LBR, MBR and HBR respectively. This gain increase shows that a better prediction is achieved by using additional information provided by the second candidate. On the JCT-VC test set, gains up to 3.3%, 2.2% and 1.7% are achieved respectively for three encoding rates.

We note that the encoding time is increased by a factor of 1.8 and 4 respectively for *SDec 1* and *SDec 2* compared with the reference HM10.1. This is mainly due to the competition of all 35 Intra directions on each candidate. The decoding time is increased by a factor of 2.6. This runtime overhead can be however reduced significantly with the use of more advanced features like parallelization techniques or shortcuts that are not within the scope of this paper.

3.3. Statistical analysis

3.3.1. SDec selection rate

Let n_{SDec} be the number of PUs encoded in SDec mode and n_{tot} be the total number of encoded PUs. We define the selection rate of SDec mode as the percentage of PUs encoded in SDec:

$$\frac{n_{SDec}}{n_{tot}} \times 100\%$$

Table 2 shows SDec selection rate of both versions *SDec 1* and *SDec 2* tested in MBR. Significant selection rates of 6.8% and 7.8% are obtained respectively, proving that proposed SDec mode can efficiently compete with existing coding modes.

Sequences class	*SDec 1*	*SDec 2*
Class A	11.4	12.8
Class B	6.7	7.8
Class C	6.6	8.0
Class D	4.6	5.4
Class E	2.5	2.7
Class F	8.9	9.9
Average	**6.8**	**7.8**

Table 2. Selection rate of SDec mode for *SDec 1* and *SDec 2* in MBR

PU size	Skip	Merge	Inter	Intra
8×8	21.3	12.4	17.7	48.5
16×16	10.6	9.9	12.4	67.1
32×32	11.2	4.9	11.6	72.3
64×64	0.1	2.2	4.2	93.5

Table 3. Distribution in percentage of classic coding modes replaced by SDec for *SDec 1* in MBR

3.3.2. Coding modes replaced by SDec mode

Considering the JCT-VC test set, table 3 gives, for each PU size, the average percentage distribution of existing coding modes that are replaced by the newly introduced SDec mode. In other words, it answers the question: what is the second best coding mode in terms of R-D cost when SDec is the optimal? The result shows that SDec mode replaces mostly HEVC Intra mode for every PU size, particularly for 64×64 PUs. Other classic modes were also replaced, confirming SDec as a promising candidate that is able to compete with existing ones.

Note that although the prediction is performed spatially, SDec is to be assimilated to an Inter mode given that it exploits a reference PU located in previously decoded frame.

4. CONCLUSION

In this paper, we have presented a new coding scheme that exploits the processing ability of the decoder to reduce the signaling of competing coding modes and parameters. They are calculated on the causal SDec reference PU rather than directly on the current PU, saving thus their transmission. This scheme allows consequently the integration of powerful tools that typically suffer from heavy signaling overhead.

Despite different restrictions applied in the presented implementation of the SDec scheme, it shows overall performance improvement when tested on a set of standard video sequences, with average bitrate savings of 1.3%, 0.9% and 0.5% respectively under LBR, MBR and HBR encoding rates.

As a future work, improving the selection of the SDec reference PU which provides optimal coding mode and parameter for encoding the current PU will be studied. It is also planned to add enhanced coding modes that will compete with current Intra mode during the SDec process.

5. REFERENCES

[1] ITU-T, "High efficiency video coding," *ITU-T Recommendation H.265*, April 2013, http://www.itu.int/rec/T-REC-H.265-201304-I.

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

[3] G. Laroche, J. Jung, and B. Pesquet-Popescu, "Intra coding with prediction mode information inference," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1786–1796, 2010.

[4] D.-Y. Kim, K.-H. Han, and Y.-L. Lee, "Adaptive intra mode bit skip in intra coding," *IEEE Asia Pacific Conference on Circuits and Systems*, pp. 446–449, 2008.

[5] S.A. Negusse, "Improving intra pixel prediction for H. 264 video coding," M.S. thesis, Blekinge Institute of Technology, May 2008.

[6] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato, and C.S. Boon, "Inter frame coding with template matching spatio-temporal prediction," *International Conference on Image Processing*, vol. 1, pp. 465–468, 2004.

[7] S. Kamp, M. Evertz, and M. Wien, "Decoder side motion vector derivation for inter frame video coding," *15th IEEE International Conference on Image Processing*, pp. 1120–1123, 2008.

[8] S. Kamp, J. Ballé, and M. Wien, "Multihypothesis prediction using decoder side motion vector derivation in inter frame video coding," *Proc. SPIE Visual Communications and Image Processing VCIP*, vol. 9, 2009.

[9] T.K. Tan, C.S. Boon, and Y. Suzuki, "Intra prediction by template matching," *IEEE International Conference on Image Processing*, pp. 1693–1696, 2006.

[10] T.T. Do, X. Lu, and J. Sole, "Compressive sensing with adaptive pixel domain reconstruction for block-based video coding," *17th IEEE International Conference on Image Processing*, pp. 3377–3380, 2010.

[11] X. Cao, C. Lai, Y. Wang, L. Liu, J. Zheng, and Y. He, "Short distance intra coding scheme for high efficiency video coding," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 790–801, 2013.

[12] J.-M. Thiesse, J. Jung, and M. Antonini, "Hybrid-1D macroblock prediction for video compression," *17th European Signal Processing Conference*, pp. 554–558, 2009.

[13] O. D. Escoda, P. Yin, C. Dai, and X. Li, "Geometry-adaptive block partitioning for video coding," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. I–657–I–660, 2007.

[14] G. Bjøntegaard, "Calculation of average PSNR differences between RD curves," *VCEG-M33, ITU-T VCEG, 13th Meeting, Austin, TX, USA*, April 2001.