INTRA RESIDUAL PREDICTION IN HEVC BY MODE DEPENDENT TEMPLATE MATCHING AND MODE DEPENDENT VECTOR QUANTIZATION

B. Huang[1,2], C. Guillemot[1], F. Henry[2], P. Salembier[3] et G. Clare[2]

[1]INRIA
[2]Orange Labs
[3]Universitat Politecnica de Catalunya

Résumé
La Norme de codage vidéo HEVC définit au total de 35 *modes de prédiction intra image qui visent à réduire la redondance spatiale en exploitant la corrélation dans les pixels à partir des références voisinages. Toutefois, dans cet article, nous montrons tout d'abord qu'il reste des corrélations non-locales après la prédiction intra image, conduisant à des résidus en haute énergie. Nous proposons ensuite deux schémas pour réduire la redondance dans le domaine residuel : l'intra prédiction résiduelle et le codage de résidu par la quantification vectorielle. Les résultats expérimentaux montrent que la prédiction de second ordre atteint* $0,2\%$ *de réduction de débit par rapport au modèle de test HEVC* 8.0. *Le codage de résidu par la quantification vectorielle améliore la performance de* 1.1% *en moyenne.*
The High Efficiency Video Coding standard (HEVC) supports a total of 35 Intra prediction modes which aim at reducing spatial redundancy by exploiting pixel correlation within a local neighborhood. However, in this paper, we first show that non-local correlation remains in the residual signals of intra prediction, leading to some high energy prediction residues. We then propose two strategies to reduce the remaining redundancy in the residual prediction domain : Intra Residual Prediction (IRP) by template matching, and residual coding using Mode Dependent Vector Quantization (MDVQ). Experimental results show that the proposed intra residual prediction achieves 0.2% *bitrate reduction at high bit rate compared with the HEVC Test Model* 8.0. *And the MDVQ-based residual coding improves the performance to* 1.1% *bitrate reduction on average.*

Mots clé : HEVC, intra prediction, residual coding, template matching, vector quantization

1. Introduction

The High Efficiency Video Coding (HEVC) standard [SOHW12] developed by the Joint Collaborative Team on Video Coding has been ratified as an international Video Coding Standard in 2013. Compared to the previous H.264/AVC standard, HEVC improves the coding efficiency of Intra prediction by introducing a larger number of prediction modes used together with a quad-tree based flexible block partitioning structure. The quad-tree based partitioning allows the splitting of a Large Coding Unit (LCU) into smaller Coding Units (CU) to represent complex structures. Up to 35 intra prediction modes are supported in HEVC to produce a more accurate prediction for smooth regions as well as directional structures. The conventional intra prediction in HEVC is efficient at reducing the local spatial redundancy in the original pixel domain. However, the accuracy of intra prediction is limited in regions having complex textures or structures. So the residues obtained by intra prediction in these regions have larger magnitudes than residues in homogeneous areas.

Different methods have been proposed in the literature for improving intra prediction, essentially for H.264/AVC. These methods can be divided into two main categories. The methods in the first category, such as Template Matching [LXWS10] and Epitome-based image compression [CGT*11], improve the coding efficiency by exploiting the non-local similarity within an image, whereas the methods in the second category, such as Second Order of Prediction (SOP) [CWLY08] [KHS09], try to reduce the correlation between the residual signals of inter prediction and the surrounding reference pixels.

In this paper, we focus on the remaining redundancy in the residual prediction domain, which we try to remove by exploiting local and non-local correlation in residual signals. Two strategies are proposed : the Intra Residual Prediction (IRP) by template matching, and the residual coding using Mode Dependent Vector Quantization (MDVQ). The proposed method of intra residual prediction exploits similarities between residual blocks resulting from intra prediction in HEVC by reusing reconstructed residual blocks. Essentially, a subset of previously decoded residue blocks is used to predict the current original residue block. Template matching helps identifying the most promising candidates within the subset. The proposed method of MDVQ-based residual coding provides another way of reducing the remaining redundancy in residual signals. Mode-dependent codebooks are learned from a training set of residue vectors which are extracted from training video sequences. These codebooks are optimized in a rate-distortion sense and do not need to be adapted for each Quantization Parameter (QP).

The rest of the paper is organized as follows. Section 2 recalls the main principles of HEVC intra prediction and quadtree partitioning. The proposed intra residual prediction is described in Section 3. Experimental results of intra residual prediction are presented in Section 4. The approach of MDVQ-based residual coding and its performance are briefly discribed in Section 5. We finally conclude this paper and discuss potential directions for future research in Section 6.

2. INTRA PREDICTION AND QUAD-TREE PARTITIONING

In HEVC, an intra predicted CU can have one or four PUs, each of which specifies a region with an individual intra prediction mode. For each PU, the best intra prediction mode is selected under a mechanism of Most Probable Mode [LBH*12]. The CU is further split into a quad-tree of Transform Units (TU), on which transform, scalar quantization and entropy coding are performed. Table 1 represents the supported CU size in HEVC and the corresponding TU size inside a CU [CGT*11].

CU size	TU size
64×64	32×32, 16×16, 8×8
32×32	32×32, 16×16, 8×8
16×16	16×16, 8×8, 4×4
8×8	8×8, 4×4

Table 1: *Supported CU size in HEVC, and available TU size inside a CU*

As mentioned in the introduction, there are remaining redundancies in the residual prediction domain. It is interesting to know the number of bits per residual-pixel required on average to code a TU. Statistical results for sequences in class D are shown in Table 2.

QP	TU 4×4	TU 8×8	TU 16×16	TU 32×32
22	1.51	0.79	0.48	1.21
27	1.07	0.47	0.31	0.54
32	0.74	0.17	0.09	0.05
37	0.43	0.17	0.09	0.05

Table 2: *Number of bits used per residual signal for a TU block*

We noticed that larger TU blocks need fewer bits per pixel than 4×4 TU blocks, so they are more efficient in the sense of bit cost. That means a 4×4 TU block has more remaining redundancy than larger TU blocks after intra prediction.

3. INTRA RESIDUAL PREDICTION BY MODE DEPENDENT TEMPLATE MATCHING

The proposed intra residual prediction by mode-dependent template matching further reduce spatial redundancy within a picture by exploiting self-similarities between residual signals of intra prediction. The scheme of intra residual prediction is shown in Figure 1. An image block is first intra predicted, which can be refered to as First Order Prediction (FOP), and the first order residue of intra prediction $\mathbf{r}_f$ is computed. The proposed intra residual prediction, which can be seen as a Second Order Prediction (SOP) is then performed on the first order residual signals. The SOP reuses a decoded first order residual block in the causal area as a predictor. The best candidate in the causal area is selected by applying a Template Matching algorithm, which is described in Section 3.3. The prediction error of the SOP, named as second order residue $\mathbf{r}_s$, is pocessed by the usual series of operations in HEVC : transform, scalar quantization and entropy coding.

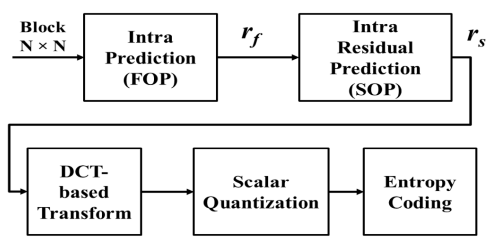

Figure 1: *Scheme of Intra Residual Prediction*

3.1. Predictor Candidate Search Region

In our method, the intra residual prediction is performed solely for 4×4 TU blocks. We reuse the reconstructed residual blocks which are aligned on a 4×4 grid as the predictor candidates. As shown in Figure 2, the search region of prediction candidate covers a causal area of the current original residual block. The size of the search region is defined by a width of $M \times 4$ and a height of $N \times 4$. Thus, a total of $M \times N$ predictor candidates are available in this search region.

3.2. Shape of Template for a Residual Block

As shown in Figure 2, the template of the original residual block to be predicted refers to the surrounding residues adjacent to the current block. Using the same principle, we define the corresponding residues in the candidate's neighborhood as the predictor template. An effective template could represent the texture or structure of the corresponding block. As a consequence, the shape of the template should adapt to the intra prediction mode of the residual block. Intra prediction modes of DC and Planar are frequently used for regions that have homogeneous texture. As shown in Figure 3(a), we use the L-formed template for these two modes. Horizontal and vertical intra prediction modes use reference pixels from the left or from above the current block respectively, and here we apply the templates in Figure 3(b) and 3(c). For diagonal modes, the shape of template depends on the directional structures of the current block, such as (d) (e) in Figure 3.

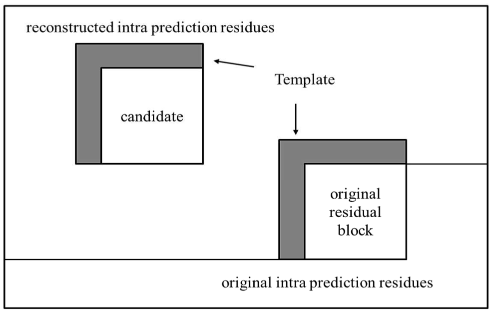

Figure 2: *Shape of template for residue block and candidate*

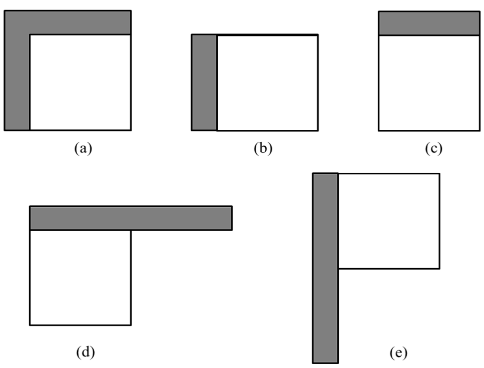

Figure 3: *Shape of mode dependent template*

3.3. Algorithm of Template Matching and Candidate List Construction

Let us denote $\mathbf{r}_f$ the original first order residual signals in the current block to be predicted by intra residual prediction. Its template containing reconstructed first order residues is denoted by $\mathbf{t}_f$. Let $\mathbf{r}_c(x,y)$ represents the reconstructed first order residues in the candidate block, which have the left upper corner residual at the position (x, y). The reconstructed first order residues in the template of this candidate is denoted as $\mathbf{t}_c(x,y)$. Then, the best candidate c^* for block $\mathbf{r}_f$ is selected from all available candidates in the search region by applying the following steps.

In step 1, the validity of a candidate is checked by using Eq.(1), which measures a square sum of all residues in the template of candidate. Candidates which do not satisfy this condition will be rejected.

$$||\mathbf{t}_c(x,y)||^2 \neq 0 \tag{1}$$

In step 2, for all available candidates, the template matching is performed by computing the Sum of Squared Error (SSE) between the residual signals in the template of the current block $\mathbf{t}_f$ and those in the template of a candidate $\mathbf{t}_c(x,y)$. Then all candidates are sorted by SSE in an ascending order.

$$SSE(\mathbf{t}_f, \mathbf{t}_c(x,y)) = ||\mathbf{t}_f - \mathbf{t}_c(x,y)||^2 \tag{2}$$

In step 3, a candidate list $\mathcal{L}$ is constructed which contains only the first N candidates in terms of lowest SSE computed by Eq.(2). A candidate is identified by its index in this list.

In step 4, the best candidate c^* in the candidate list $\mathcal{L}$ is the one containing reconstructed residues $\mathbf{r}_{c^*}$ that is most similar to the original residual singals $\mathbf{r}_f$ in the current block to be predicted. This can be expressed as :

$$c^* = \arg\min_{c \in \mathcal{L}} ||\mathbf{r}_f - \mathbf{r}_c(x,y)||^2 \tag{3}$$

The difference between the reconsturcted first order residues in the best candidate residual block and the original first order residues in the current block is refered to as second order residue $\mathbf{r}_s$. It should be noted that, the intra residual prediction (SOP) is performed only if it reduces the energy of the first order residue. Otherwise, the intra residual prediction is skipped. This condition can be expressed as :

$$||\mathbf{r}_f - \mathbf{r}_{c^*}(x,y)||^2 < ||\mathbf{r}_f||^2 \tag{4}$$

3.4. Structure of Signaling and Decoding Scheme

By analyzing the position of the 4×4 TU blocks in which the intra residual prediction is activated, we found that for some CUs, the nested TUs tend to have the same decision on the use of intra residual prediction. Consequently, we use three syntax elements and a tree structre for the signalization of the proposed intra residual prediction. For a CU block containing 4×4 TU blocks, the syntax element irp_cu_flag indicates whether one of the nested TUs on which the IRP is performed. This flag is coded using one bit and an associated context-adaptive binary arithmetic coding (CABAC) context. When irp_cu_flag is set to 1, for each 4×4 TU block inside this CU, a syntax element irp_tu_flag is signalled to indicate whether the IRP is performed. This syntax element is coded using one bit and an associated CABAC context. When the IRP is used, a fixed-length syntax element irp_tu_idx is transmitted. It represents the index of the best predictor candidate in the candidate list.

At the decoder side, the syntax element irp_cu_flag of a CU block is parsed firstly. If the flag value is zero, none of the inside TUs on which the intra residual prediction is performed. Otherwise, the syntax element irp_tu_flag is then parsed for each 4×4 TU block. Finally, the index of the best candidate is parsed for TU block having syntax element *irp_tu_flag* equals 1. A candidate list is constructed using the procedure described in Section3.3. The decoded residues of the current block actually correspond to the second order residues of the intra residual prediction. By adding the predictor identified by the decoded index to this second order residue, the first order residue is reconstructed.

4. EXPERIMENTAL RESULTS OF INTRA RESIDUAL PREDICTION

The proposed intra residual prediction was implemented in the reference software HM8.0 [C7] of the HEVC test Model. Since our algorithm aims at improving the intra coding efficiency, all the frames of a sequence were intra coded. The encoder is configured following the JCT-VC Common Test set with the All-Intra profile. The performance is evaluated by comparing our algorithm with HM8.0. Eight video sequences (ParkScence, Cactus, BasketballDrill, BQMall, BasketballPass, BlowingBubbles, BQSquare, RaceHorses) with different resolutions as specified in HEVC [F.B11] were encoded. The experiments are performed under two ranges of quantization parameters : mid-bitrate (MBR) for QP values 22, 27, 32 and 37 ; and high-bitrate (HBR) for QP values 16, 21, 26, 31. The BD-Rate performances of the proposed algorithm are measured with the method in [Bjø01]. Table 3 shows the performance where the negative values mean a bitrate saving.

Sequences	**HBR** (%)	**MBG** (%)
ParkScene	−0.06	−0.02
Cactus	−0.07	−0.04
BasketballDrill	−0.18	−0.12
BQMall	−0.11	−0.06
BasketballPass	−0.09	−0.04
BlowingBubbles	−0.08	−0.04
BQSquare	−0.07	−0.04
RaceHorse	−0.07	−0.01

Table 3: *Simulation results in iterms of bitrate saving*

From the results, we can see that the proposed algorithm provides a bit rate saving of up to 0.18% under high-bitrate compared with HM8.0. It should be noted that, without the cost of signalization in the intra residual prediction, a bitrate reduction of 13% could be achieved, a theoretical upper limit that gives an indication about the possible improvements. We conclude that the template matching is not the most efficient method for identifying the best candidate of intra residual prediction, due to higher cost of signalling the predictor.

5. MDVQ-BASED RESIDUAL CODING

The approach of residual coding using MDVQ provides another way of exploiting the correlations in residual signals. Actually, in a block of samples intra predicted with a planar mode or DC mode, the residual signals have a relatively homogeneous structure, whereas those derived from angular prediction modes tend to have directional structures. In our approach, the codebooks of vector quantization are learned with the aim of modelling the directional characteristics of the intra prediction residual signals.

The scheme of MDVQ-based residual coding is similar to the intra residual prediction shown in Figure 1, whereas the IRP is replaced by the MDVQ-based residual coding. For a image block which is intra predicted with mode i, the first order original residue of intra prediction $\mathbf{r}_f$ is then quantized by a matching codevector $\mathbf{v}$ in a pre-generated codebook C_i. The vector quantization error (the difference between the original residue and the matching codevector) which can be refered to as the second order residue $\mathbf{r}_s$, is processed by the following operations : transform, scalar quantization and entropy coding. In our experiments, the MDVQ-based residual coding has been used for the TU block size of 4×4 and 8×8. For each of them, 35 codebooks corresponding to the 35 intra prediction modes are derived. Each codebook contains 256 codevectors, which represents a good compromise between storage requirements, complexity and performance. The codebook learning processe is iterated, based on a training set of original residue vectors. They are obtained by extracting from training video sequences those residue vectors selected by the Lagrange Rate Distortion Optimization.

The experiments are performed under the mid-bitrate range of quantization parameter. The encoder is configured as the experiments of intra redisual prediction described is Section 4. Table 4 shows the performance of MDVQ-based residual coding of sequences of six test classes. Here the MDVQ codebooks are trained on sequences that are different from those used to measure the performance. This is the realistic use case. One can observe an average bitrate saving of 1.1%. Interestingly, a larger gain is observed for low-resolution sequences which are usually more difficult to compress, this being attributed to the larger proportion of 4×4 and 8×8 TU blocks. The method also performs well on videoconference (class E) and screen content (class F) sequences.

By utilizing codebooks which are well suited to the test sequences, such as the training sequence for codebook generation is the same as the test sequences, better performance can be achieved. Table 5 shows the experimental results of MDVQ-based residual coding on class B sequences. One can observe that our method provides a bitrate reduction of 4.9% on average in this ideal-but not realistic-case. This experiments demonstrate that there are still rooms for improving the MDVQ-based residual coding.

6. CONCLUSION

The proposed intra residual prediction scheme aims to exploit non-local correlation in the residual domain within a picture. By making use of the adequate shape of template in template matching, intra residual prediction can further reduce the redundancy of residual signal of the conventional prediction approaches. The MDVQ-based residual coding provides a better scheme to reduce the remaining redundancy in residual signals, while further tests indicate that codebook adaptivity could substantially improve the performance.

sequence class	bit-rate saving (%)
NebutaFestival	−0.09
PeopleOnStreet	−1.62
SteamLocomotiveTrain	0.44
Traffic	−1.51
average of class A	−0.7
BasketballDrive	0.12
BQTerrace	−1.33
Cactus	−1.03
Kimono	0.02
ParkScene	−1.44
average of class B	−0.5
BasketballDrill	−2.27
BQMall	−1.34
ParkScene	−0.95
RaceHorse	−0.52
average of class C	−1.3
BasketballPass	−1.26
BlowingBubbles	−1.24
BQSquare	−0.98
RaceHorse	−1.68
average of class D	−1.3
FourPeople	−1.61
Johnny	−1.67
KristenAndSara	−1.2
average of class E	−1.6
BasketballDrillText	−2.53
ChinaSpeed	−1.29
SlideEditing	−0.91
SlideShow	−1.80
average of class F	−1.6

Table 4: *Bitrate savings using sequence-independent codebooks*

sequence name	bit-rate saving (%)
Kimono1_1920x1080p	−0.25
ParkScene_1920x1080p	−3.19
Cactus_1920x1080p	−13.66
BQTerrace_1920x1080p	−1.99
BasketballDrive_1920x1080p	−5.16

Table 5: *Bitrate savings using sequence-dependent codebooks*

Références

[Bjø01] BJØNTEGAARD G. : Calculation of average psnr differences between rd-curves. *ITU-T SG16 Q.6 Doc. VCEG-M33* (avril 2001).

[C7] Hevc test model (online). `http://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware`.

[CGT*11] CHERIGUI S., GUILLEMOT C., THOREAU D., GUILLOTEL P., PEREZ P. : Epitome-based image compression using translational sub-pel mapping. *MMSP* (octobre 2011).

[CWLY08] CHEN S., WANG J., LI S., YU L. : Second order prediction (sop) in p slice. *ITU-T SG16/Q.6 Doc. VCEG-AI27* (juillet 2008).

[F.B11] F.BOSSEN : Common test conditions and software reference configurations. *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T VCEG and ISO/IEC MPEG. Doc. JCTVC-G1200* (février 2011).

[KHS09] KIM S., HWANG S., SUNWOO M. : Novel residual prediction scheme for hybrid video coding. *ICIP* (novembre 2009).

[LBH*12] LAINEMA J., BOSSEN F., HAN W., MIN J., UGUR K. : Intra coding of the hevc standard. *IEEE Transaction on Circuits and Systems for Video Technology. Vol. 22*, Num. 12 (décembre 2012).

[LXWS10] LAN C., XU J., WU F., SHI G. : Intra frame coding with template matching prediction and adaptive transform. *ICIP* (septembre 2010).

[SOHW12] SULLIVAN G., OHM J., HAN W., WIEGAND T. : Overview of the high efficiency video coding (hevc) standard. *IEEE Transaction on Circuits and Systems for Video Technology* (décembre 2012).