TWO-STEP PROGRESSIVE INTRA PREDICTION FOR VERSATILE VIDEO CODING

Meng Lei[1], *Falei Luo*[1], *Xinfeng Zhang*[2], *Shanshe Wang*[1], *Siwei Ma*[1]

[1]Institute of Digital Media, Peking University, Beijing China,
[2]University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

In traditional intra prediction, nearest reference samples are utilized to generate the prediction block. Although more directional intra modes and reference lines have been utilized, encoders could not predict complex content with only the local reference samples efficiently. To address this issue, a two-step progressive prediction method combining local and non-local information is proposed. The non-local information can be obtained through template matching based prediction, and the local information can be derived by the high frequency coefficients from the first prediction step. Experimental results show that the proposed method can achieve 0.87% BD-rate reduction in VTM-7.0. In particular, the method is of significant advantages over prediction schemes using only non-local information.

Index Terms— Video coding, VVC, Intra prediction, template matching, low frequency coefficients

1. INTRODUCTION

Demands for high resolution and high quality video has promoted fast development of video coding standards. In 2015, the committee Joint Video Exploration Team (JVET), making up of experts in ISO Motion Picture Expert Group and ITU-T Video Coding Expert Group, started to develop the next-generation video coding standard. Two years later, the exploration platform of JVET had obtained 30% coding efficiency improvement over High Efficiency Video Coding (HEVC). Then a joint call for proposals was issued to solicit proposals for the new generation video coding standard, named Versatile Video Coding (VVC). Till now, VVC has achieved nearly 35% bit-rate saving compared to HEVC [1, 2].

In current design, intra prediction is conducted by filtering the neighbouring reconstructed samples along a certain direction. In this way, spatial correlation in video content is utilized to improve the coding efficiency. Furthermore, more flexible block partitions, prediction modes and reference lines help improve the coding gain [3].

In VVC, a quadtree (QT) with nested multi-type tree (MTT) using binary and ternary splits segmentation structure replaces the concepts of multiple partition unit types. In the MTT structure, both binary-tree (BT) partition and ternary-tree (TT) partition can be performed recursively after the QT, and there are horizontal and vertical splitting directions in BT and TT partitions [4]. In particular, for intra-coded blocks, intra sub-partitions (ISP) is proposed to divide luma intra-predicted blocks vertically or horizontally into 2 or 4 sub-partitions depending on the block size [5]. Thus short-distance prediction can be achieved to obtain more accurate prediction samples.

The number of directional intra modes in VVC has been extended from 33 to 65 to capture the arbitrary edge directions, and the planar and DC modes remain the same [6]. Meanwhile, in order to accommodate the the non-square blocks, several conventional angular intra prediction modes are adaptively replaced with wide-angle intra prediction modes [7, 8]. Considering that the nearest reference line may have noise, multiple reference line (MRL) is also utilized for intra prediction in VVC, in which 2 additional reference lines are used for prediction [9]. Matrix weighted intra prediction (MIP) method is a newly added intra prediction technique into VVC. MIP takes the left and above neighbouring samples for prediction using linear averaging [10]. Based on the structure, more neighbouring pixels can be combined sufficiently to obtain better prediction samples. Position dependent intra prediction combination (PDPC) is an intra prediction method to enhance the prediction quality of the planar mode, in which the results of conventional intra prediction are further modified by a weighted filtering process and the weighting is related to the position of prediction sample [11]. Additionally, a cross-component linear model (CCLM) was proposed to reduce the cross-component redundancy by utilizing the previously decoded structure of luma samples [12, 13].

These tools can make advantage of the nearest neighbouring reconstructed samples and local correlations to improve the coding efficiency. However, intra prediction in VVC still cannot deal with complicated textures well due to the limitation of local reference samples. In this paper, we focus on combining non-local and local correlations to further decrease the residuals to improve the intra coding performance.

This work was supported by Key-Area Research and Development Program of Guangdong Province (2019B010133001), National Key Research and Development Project (2019YFF0302703) and High-performance Computing Platform of Peking University, which are gratefully acknowledged.

978-1-7281-6395-6/20/$31.00 ©2020 IEEE

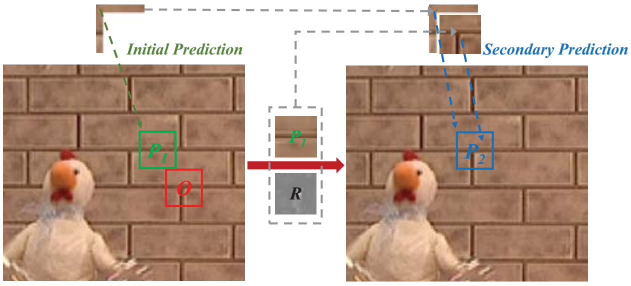

Fig. 1. Illustration of two-step progressive intra prediction.

2. MOTIVATION

Although intra prediction in VVC is more precise, many bits are still needed to signal quantized residual coefficients. The low frequency coefficients tend to be larger and consume more bits to signal, conceivably the coding efficiency would improve if an accurate estimation of low frequency coefficients could be achieved.

In [14], a DC coefficient estimation algorithm was proposed based on the border continuities. Discarding the DC coefficient directly in each transform block produce strong discontinuities between neighbouring blocks. The algorithm aimed to solve an optimal offset to recover the corresponding block edges by minimizing the sum of gradient along the prediction direction. However, the DC coefficient was difficult to represent with a single offset as more transform types and non-separable secondary transform introduced in VVC, and it is not completely reasonable to minimize the gradient along the prediction direction to solve the offset. Motivated by that the samples reconstructed with AC coefficients can reflect the general texture in [14], we reconstruct high frequency coefficients as local information to predict low frequency coefficients.

Considering that the utilization of local information is limited in recovering low frequency information, the non-local search algorithm is introduced to achieve efficient prediction. Template matching (TM) is a potential technique which searches a similar non-local block by using neighbouring reconstructed samples as template [15]. Due to the high decoding complexity, many works focus on reducing the complexity increase from the implied search process [16, 17]. Intra block copy (IBC) is a similar technique in which the predictor is indicated by a block vector (BV) [18]. Compared with TM, IBC can find more accurate prediction blocks but also requires more bits to express prediction information.

Combining the neighbouring reconstruction samples and local samples reconstructed with high frequency coefficients as a template, a more accurate non-local similar block could be searched without the block vector signalling. Hence the low frequency coefficients could be predicted with the non-local similar blocks.

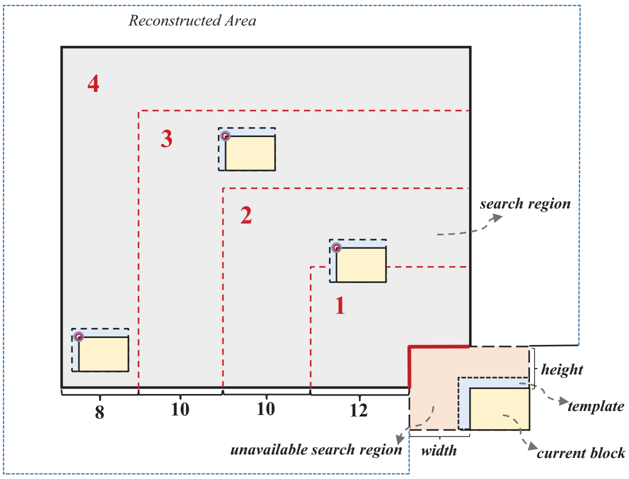

Fig. 2. Search template and four regions of TMP.

3. PROPOSED METHOD

In VVC, only the neighbouring samples were utilized in intra prediction. In this way, the spatial redundancy could not be fully eliminated. In this section, we proposed a two-step progressive intra prediction algorithm, aiming to improve the prediction accuracy by combining the local and non-local content similarity to derive better prediction results.

The framework of the proposed method is shown in Fig 1. Firstly template matching based prediction (TMP) is performed to derive the preliminary prediction block. Then the residual coefficients in the first and second frequency bands in scanning order are set to be 0, and dequantization and inverse transform on the remaining coefficients are applied. Thus a reconstruction block is generated. Later, the reconstruction block and adjacent reference samples are combined together as a template to perform the TMP again. The results of the secondary prediction is regarded as the actual prediction of current block and the coefficients of the first and second frequency bands will be updated. To enable the proposed feature, a flag indicating whether this mode is utilized should be transmitted for each PU. In the following sub-sections, we give the details of the two steps for this algorithm.

3.1. Template matching based prediction

Template matching (TM) is a technique that uses a template image to find the most similar structure in a reference picture. In order to get the samples reconstructed with partial transform coefficients to assist subsequent prediction, we adopt template matching based method to obtain preliminary prediction block. In this paper, the top reference samples, top-left reference sample and left border samples make up the searching template.

Since there are a large quantity of candidate positions to

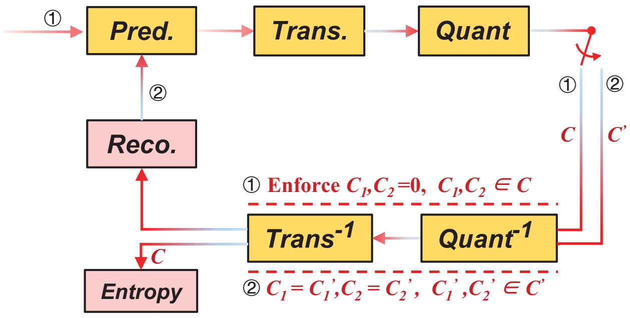
①
Pred.
Trans.
Quant
②
Reco.
① ②
C C'
① Enforce $C_1,C_2=0,\ C_1,C_2 \in C$
Trans^{-1}
Quant^{-1}
C
Entropy
② $C_1=C_1', C_2=C_2',\ C_1',C_2' \in C'$

Fig. 3. Workflow of progressive secondary prediction.

search for the best matchings, the computing complexity of TM is very high. To decrease the complexity, [17] proposed to restrict the searching area to a limited region and divide the region into several small parts. In this paper, we utilized the similar restriction and divide the searching region into 4 parts as shown in Fig. 2. In each search region, three best matching candidates are selected by minimizing the mean square error (MSE) between the template and the matching template. Furthermore, weighted averaging of three best matching blocks inside one region is utilized as the first pass prediction of current block, which is denoted as follows:

$$P_{S1,i} = \begin{cases} \frac{2P_{i,1}+P_{i,2}+P_{i,3}}{4}, & \mathrm{if} E_{i,2}<2E_{i,1} \& E_{i,3}<2E_{i,1} \\ \frac{P_{i,1}+P_{i,2}}{2}, & \mathrm{elif} E_{i,2}<2E_{i,1} \\ P_{i,1}, & \mathrm{otherwise} \end{cases} \tag{1}$$

where $P_{i,1}$, $P_{i,2}$, $P_{i,3}$ is the corresponding reconstructed block associated with the three best matching position in the i-th region, and $E_{i,1}$, $E_{i,2}$, $E_{i,3}$ is their respective MSE. $P_{S1,i}$ is the prediction result of the first pass prediction, which would be utilized to predict the low frequency coefficients.

When the searching of every region is accomplished, the region with the smallest MSE between the prediction samples and the original samples will be determined as the best region. The corresponding prediction will be denoted as P_{S1}, and the index of the region with the best prediction will be transmitted through the bit stream. Then the TMP could be conducted within the selected searching region and the prediction could be obtained in a decoder.

3.2. Progressive secondary prediction

Based on the consistency of the best BV derived by the encoding block to the best BV derived by the neighbouring samples, the adjacent reconstruction samples were utilized as a matching template to find the matching block for TMP. However, in videos with complex texture, flexible content or noise, there would be much difference in the matching result relying only on L-shaped template, this will introduce large residual when TMP is used. To improve the prediction accuracy, a progressive prediction scheme is proposed, where the high frequency coefficients of the first prediction step were utilized to assist the secondary prediction as shown in Fig. 3. The proposed method is conducted as follows:

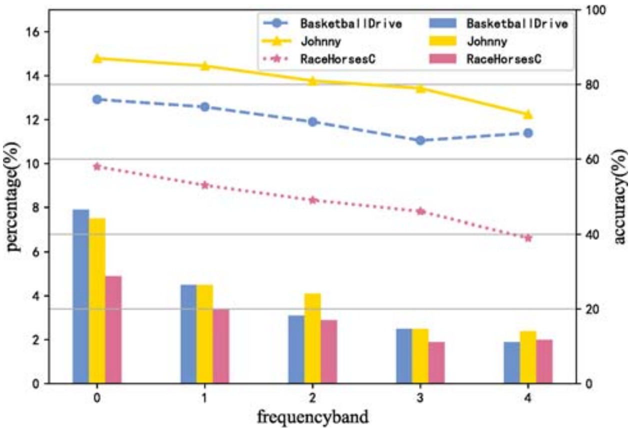

Fig. 4. The prediction accuracy and proportion of bits from coefficients in different frequency bands.

Firstly, the first pass prediction P_{S1} is conducted to derive the residual coefficients denoted as C. We subtract the input samples with the P_{S1} to obtain the residuals. Then the residual are transformed and quantized to get the quantized coefficients following the traditional process. Later, the low frequency coefficients are set to zero and the reconstruction process with the remaining coefficients is conducted. Thus, reconstruction block for the first step prediction, denoted as T_c, is generated.

Secondly, a combined template is derived to conduct TMP again. The neighbouring L-shaped reconstructed template called T_r together with T_c are combined as the final template called T_w. Then T_w is used as template to perform TMP in the search area shown in Fig. 1. To balance performance and the decoder complexity, the search region is limited to the optimal region selected in the preliminary prediction. The optimal matching position is determined by minimizing E_w, where E_w is calculated as:

$$E_w = E_r + E_c \tag{2}$$

where E_r and E_c is the MSE between matching template and T_r, T_c. Then the reconstruction samples corresponding to the optimal prediction position, denoted as P_{S2} is obtained, and it is the final prediction of current block.

Lastly, the low frequency coefficients are updated. The residual of the second pass prediction P_{S2} is transformed and quantized to get the new coefficients, which is denoted as C'. Then the low frequency bands of C' with the coefficients in C are combined together to generate the final quantized coefficients. These coefficients are transmitted to the decoders. Hence, the final reconstruction samples can be reconstructed by the new prediction samples and coefficients.

Table 1. The performance of the proposed method compared with VTM7.0

Sequence	**Over VTM7.0-AI**		
	Y	**U**	**V**
Class A1	-0.38%	-0.53%	-0.22%
Class A2	-0.81%	-0.95%	-0.97%
Class B	-0.72%	-0.67%	-0.61%
Class C	-0.85%	-0.82%	-0.77%
Class E	-1.66%	-1.50%	-1.48%
Overall	-0.87%	-0.86%	-0.79%
Class D	-0.56%	-0.71%	-0.54%
Class F	-1.31%	-0.88%	-0.67%

Table 2. Comparison with [17] on BMS-2.0.1 under AI configuration

Sequence	**L0077 [17]**			**Proposed over BMS-2.0.1**		
	Y	**U**	**V**	**Y**	**U**	**V**
Class A1	-0.24%	-0.10%	-0.12%	-0.66%	-0.56%	-0.74%
Class A2	-0.99%	-0.87%	-0.88%	-1.28%	-1.52%	-1.47%
Class B	-0.86%	-0.65%	-0.71%	-1.19%	-1.06%	-1.17%
Class C	-1.02%	-0.84%	-0.90%	-1.46%	-1.36%	-1.35%
Class E	-2.04%	-1.92%	-2.06%	-2.25%	-2.30%	-2.42%
Overall	-1.01%	-0.85%	-0.91%	-1.35%	-1.43%	-1.49%
Class D	-0.51%	-0.36%	-0.69%	-0.82%	-0.81%	-0.90%
Class F	-8.19%	-8.05%	-8.16%	-9.82%	-8.33%	-8.93%

As we can see, in the proposed secondary prediction, not only the adjacent reconstruction information but also the texture information of the current block are used, so that the prediction blocks with consistent texture trends can be matched better. In this way, the residuals are smaller, which can reduce the bits for encoding residual coefficient. And it is also possible to impair the quantization error when the result of secondary prediction is used as the final prediction samples.

Fig. 4 shows the influence of low frequency coefficients. The histograms represented the proportion of bits for encoding coefficients in different frequency bands, and the line charts represented prediction accuracy. The prediction accuracy indicates the consistency of the best matching positions to the search positions when the corresponding frequency band was enforced to be zero. To balance prediction accuracy and bit savings, only the first and second frequency bands were set to be zero in this paper.

4. EXPERIMENTAL RESULTS

The proposed algorithm is evaluated under the reference software VTM7.0 of the emerging VVC standard. The test conditions are consistent with the common test conditions [19]. In this paper, the first 200 frames of common test sequences are encoded, using the all intra (AI) configuration of VVC. Here, four common test QP values $\in \{22, 27, 32, 37\}$ were chosen to encode these clips. The detailed performance evaluation results of the proposed method under AI configuration are shown in Table 1, where the coding efficiency is measured in terms of Bjontegaard delta bitrate (BD-BR) [20].

As we can seen, the proposed algorithm achieves 0.87% improvement under AI configuration. Especially for screen content sequences (Class F), the algorithm can still achieve 1.31% BD-rate reduction on average with IBC on. Furthermore, the search results are also shown in Fig. 5, where the red, green and blue borders represent the blocks encoded by our proposed method and the blocks searched by the preliminary prediction and secondary prediction, respectively. It can be seen that the results of secondary prediction can refine the results of the preliminary prediction and find more accurate similar blocks.

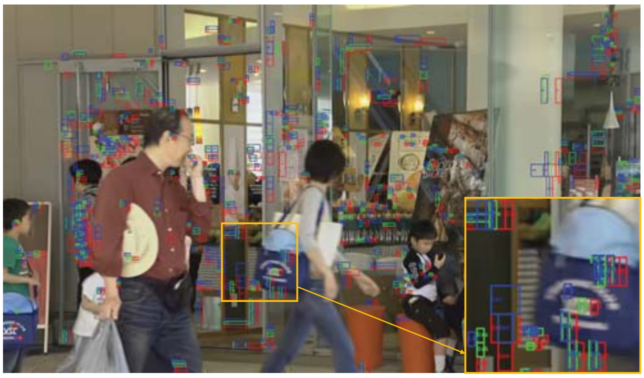

Fig. 5. The search results extracted from frame 17 of *BQMall*.

We further compare our proposed scheme with region-based TMP [17], shown in Table 2. It is obvious that our algorithm outperforms the region-based TMP especially for sequences with complex texture or noisy, e.g. *Campfire* and *RaceHorses*, which proves that our progressive secondary prediction has more accurate prediction efficiency.

5. CONCLUSION

In this paper, a two-step progressive intra prediction method is presented for VVC intra coding. The proposed scheme aims to improve prediction accuracy and reduce low-frequency coefficients using both local and non-local correlation. First the template matching based prediction is employed to obtain preliminary prediction samples and quantized coefficients. The secondary prediction is then performed to search a better prediction block by using local samples reconstructed with preliminary quantized coefficients. Thereby both prediction samples and quantized coefficients can be updated. Experimental results indicates our proposed algorithm achieves 0.87% bit-rate savings under AI configuration.

6. REFERENCES

[1] 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*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] J. Chen, M. Karczewicz, Y. Huang, K. Choi, J.-R. Ohm, and G.J. Sullivan, "The Joint Exploration Model (JEM) for Video Compression with Capability beyond HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, 2019.

[3] J. Li, B. Li, J. Xu, and R. Xiong, "Efficient Multiple-Line-Based Intra Prediction for HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 4, pp. 947–957, 2018.

[4] J. Chen, Y. Ye, and S.-H. Kim, "Algorithm description for Versatile Video Coding and Test Model 7 (VTM 7)," JVET-P2002, Joint Video Exploration Team (JVET), Oct. 2019.

[5] S. De-Luxán-Hernández, V. George, J. Ma, T. Nguyen, H. Schwarz, D. Marpe, and T. Wiegand, "An Intra Subpartition Coding Mode for VVC," *IEEE International Conference on Image Processing (ICIP)*, pp. 1203–1207, Aug. 2019.

[6] N. Choi, Y. Piao, K. Choi, and C. Kim, "CE3.3 related:Intra 67 modes coding with 3MPM," JVET-K0529, Joint Video Exploration Team (JVET), Jul. 2019.

[7] F. Racape, G. Rath, F. Urban, L. Zhao, S. Liu, X. Zhao, X. Li, A. Filippov, V. Rufitskiy, and J. Chen, "CE3-related:Wide-angle intra prediction for non-square blocks," JVET-K0500, Joint Video Exploration Team (JVET), Jul. 2019.

[8] L. Zhao, X. Zhao, S. Liu, X. Li, J. Lainema, G. Rath, and F. Urban, "Wide Angular Intra Prediction for Versatile Video Coding," *2019 Data Compression Conference (DCC)*, pp. 53–62, May. 2019.

[9] Y.-J. Chang, H.-J. Jhu, H.-Y. Jiang, L. Zhao, S. Liu, and T. Wiegand, "Multiple Reference Line Coding for Most Probable Modes in Intra Prediction," *2019 Data Compression Conference (DCC)*, p. 559, May. 2019.

[10] M. Schafer, B. Stallenberger, J. Pfaff, P. Helle, H. Schwarz, D. Marpe, and T. Wiegand, "An Affine-Linear Intra Prediction With Complexity Constraints," *IEEE International Conference on Image Processing (ICIP)*, pp. 1089–1093, Aug. 2019.

[11] A. Said, X. Zhao, M. Karczewicz, J. Chen, and F. Zou, "Position dependent prediction combination for intra-frame video coding," *IEEE International Conference on Image Processing (ICIP)*, pp. 534–538, Aug. 2016.

[12] J. Chen, "Chroma Intra Prediction by Scaled Luma Samples Using Integer Operations," JCTVC-C206, Joint Collaborative Team on Video Coding (JCT-VC), Oct. 2010.

[13] K. Zhang, J. Chen, L. Zhang, X. Li, and M. Karczewicz, "Enhanced Cross-Component Linear Model for Chroma Intra-Prediction in Video Coding," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3983–3997, 2018.

[14] C. Chen, Z. Miao, X. Meng, S. Zhu, and B. Zeng, "DC Coefficient Estimation of Intra-Predicted Residuals in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 8, pp. 1906–1919, 2018.

[15] T.K. Tan, C.S. Boon, and Y.Suzuki, "Intra prediction by template matching," *IEEE International Conference on Image Processing (ICIP)*, pp. 1693–1696, Feb. 2007.

[16] G. Venugopal, P. Merkle, D. Marpe, and T. Wiegand, "Fast template matching for intra prediction," *IEEE International Conference on Image Processing (ICIP)*, pp. 1692–1696, Feb. 2018.

[17] G. Venugopal, K. Müller, H. Schwarz, D. Marpe, and T. Wiegand, "CE8: Intra Region-based Template Matching (Test 8.1)," JVET-L0077, Joint Video Exploration Team (JVET), Oct. 2018.

[18] X. Xu, S. Liu, T. Chuang, Y. Huang, S. Lei, K. Rapaka, C. Pang, and M. Karczewicz, "Intra Block Copy in HEVC Screen Content Coding Extensions," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 409–419, 2016.

[19] A. Segall, E. François, W. Husak, S. Iwamura, and D. Rusanovskyy, "JVET common test conditions and evaluation procedures for HDR/WCG video," JVET-P2002, Joint Video Exploration Team (JVET), Oct. 2019.

[20] G. Bjontegaard, "Calculation of average psnr difference between rd-curves," 13th VCEG-M33 Meeting, Austin, TX, Apr. 2001.