

LETTER

New VVC Chroma Prediction Modes Based on Coloring with Inter-Channel Correlation

Zhi LIU^{†a)}, Member, Jia CAO[†], Xiaohan GUAN[†], and Mengmeng ZHANG^{†,††b)}, Nonmembers

SUMMARY Inter-channel correlation is one of the redundancy which need to be eliminated in video coding. In the latest video coding standard H.266/VVC, the DM (Direct Mode) and CCLM (Cross-component Linear Model) modes have been introduced to reduce the similarity between luminance and chroma. However, inter-channel correlation is still observed. In this paper, a new inter-channel prediction algorithm is proposed, which utilizes coloring principle to predict chroma pixels. From the coloring perspective, for most natural content video frames, the three components Y, U and V always demonstrate similar coloring pattern. Therefore, the U and V components can be predicted using the coloring pattern of the Y component. In the proposed algorithm, correlation coefficients are obtained in a lightweight way to describe the coloring relationship between current pixel and reference pixel in Y component, and used to predict chroma pixels. The optimal position for the reference samples is also designed. Based on the selected position of the reference samples, two new chroma prediction modes are defined. Experiment results show that, compared with VTM 12.1, the proposed algorithm has an average of -0.92% and -0.96% BD-rate improvement for U and V components, for All Intra (AI) configurations. At the same time, the increased encoding time and decoding time can be ignored.

key words: chroma prediction, coloring pattern, inter-channel correlation, VVC

1. Introduction

Versatile Video Coding (VVC) [1] is the newest generation video compression standard finalized by the Joint Video Experts Team (JVET) in July 2020. The VVC standard demonstrates at least 40% BD-rate savings than its previous generation video compression HEVC/H.265 on Ultra High Definition test content.

In VVC, a series of new intra prediction tools have been introduced, including the Intra Sub-partition (ISP) [2], Matrix-based Intra Prediction (MIP) [3], Multiple Reference Line Prediction (MRLP) [4] and Wide-Angle Intra Prediction (WAIP) [5], and so on. The Cross-Component Linear Model (CCLM) tool that exploits the inter-channel redundancy is adopted for chroma intra prediction in VVC [6]. It predicts the chroma samples of the block based on the linear model of the reconstructed samples in the luminance block of the same position. Up to now, there are several improve-

ments for CCLM, including Multi-Model Linear Model (MMLM) [7], Multi-Filter Linear Model (MFLM) [7], and Multi-Directional Linear Model (MDLM) [8]. All of them can achieve significant coding gains.

Meanwhile, to further exploit the efficient representation from luminance to chroma, the deep learning networks have been adopted. Blanch et al. [9] presented a neural network architecture for cross component intra prediction, in which an attention module was employed for learning spatial relations. In [10], the authors further improved the model and reduce the complexity. Zhu et al. [11] presented a new neural network architecture for chroma prediction. These methods can significantly improve the coding efficiency, however, the computation complexity is increased dramatically.

In this paper, a new idea based on coloring principle is introduced to design lightweight chroma prediction algorithm. The main idea is that, for most natural content video frames, the three components Y, U and V always demonstrate similar coloring pattern. Therefore, the U and V components can be predicted using the coloring pattern of the Y component. The experiments illustrated that the proposed method can improve the compression efficiency of chroma components by more than 0.9% in All Intra (AI), while the increasing of encoding time and decoding time can be ignored.

The remainder of the paper is organized as follows. A brief description of Chroma Prediction Modes in VVC is provided in Sect. 2. In Sect. 3, we describe the proposed method. The performance of the proposed method is presented in Sect. 4. Finally, Sect. 5 provides the conclusion.

2. VVC Chroma Prediction Modes

In CCLM mode, the relationship between luminance and chroma is described as:

$$Predc(i, j) = \alpha * Rec'L(i, j) + \beta \quad (1)$$

where $Predc(i, j)$ represents a predicted chroma sample at position (i, j) in a coding unit and $Rec'L(i, j)$ represents its corresponding down-sampled reconstructed luminance sample in the same coding unit.

The VVC standard takes a low-complexity max-min method to derive the parameters α and β in the linear model. At most four neighboring chroma samples (C_1, C_2, C_3, C_4) and their corresponding down-sampled luminance samples (Y_1, Y_2, Y_3, Y_4) at pre-defined positions are selected and

Manuscript received February 3, 2022.

Manuscript revised April 27, 2022.

Manuscript publicized June 27, 2022.

[†]The authors are with North China University of Technology, Beijing, China.

^{††}The author is with Beijing Polytechnic College, Beijing, China.

a) E-mail: lzliu@ncut.edu.cn (Corresponding author)

b) E-mail: muchmeng@126.com (Corresponding author)

DOI: 10.1587/transinf.2022EDL8010

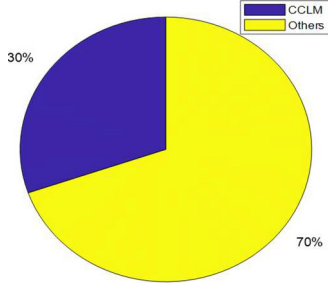


Fig. 1 Ratio of selection CCLM in class D

averaged to form the two luminance and chroma sample pairs, denoted as Y_a , Y_b and C_a , C_b . α and β are derived as follows:

$$\begin{cases} C_a = (C_1 + C_2 + 1) \gg 1; \\ C_b = (C_3 + C_4 + 1) \gg 1; \\ Y_a = (Y_1 + Y_2 + 1) \gg 1; \\ Y_b = (Y_3 + Y_4 + 1) \gg 1; \\ \alpha = \frac{Y_b - Y_a}{C_b - C_a}; \\ \beta = Y_b - \alpha * C_b; \end{cases} \quad (2)$$

In this study, the proportion of choosing CCLM as the final optimal mode in intra chroma prediction stage by the encoder for D class sequences is counted as Fig. 1. As can be found from the figure, this proportion is about 30%, indicating that there are a large number of the pixels that are not meet the linear modal defined by CCLM. For these kind of pixels, other efficient models should be found.

3. Proposed Algorithm

3.1 Main Idea

In digital image processing field, colorization has been widely accepted as a technology to convert grayscale images to color ones. Local color expansion [12] is one of the typical colorization methods worked in YUV color space. It output the resulting $U(r)$, $V(r)$ from the input luminance Y , by making use of the constraint that two neighboring chroma pixels (assumed to be r , s) should have similar coloring pattern as their corresponding luminance pixels. It tries to minimize the difference between the chroma $U(r)$ of pixel r and its weighted average of the neighboring pixels [12]:

$$J(U) = \sum_r (U(r) - \sum_{s \in N(r)} W_{rs} \cdot U(s))^2 \quad (3)$$

where $s \in N(r)$ denotes the fact that r and s are neighboring pixels, and W_{rs} is based on the normalized correlation between the two intensities:

$$W_{rs} \propto 1 + \frac{1}{\sigma_r^2} (Y(r) - u_r)(Y(s) - u_r) \quad (4)$$

where u_r and σ_r are the mean and variance of the luminance in a window around r . From (4), it can be found that W_{rs} is

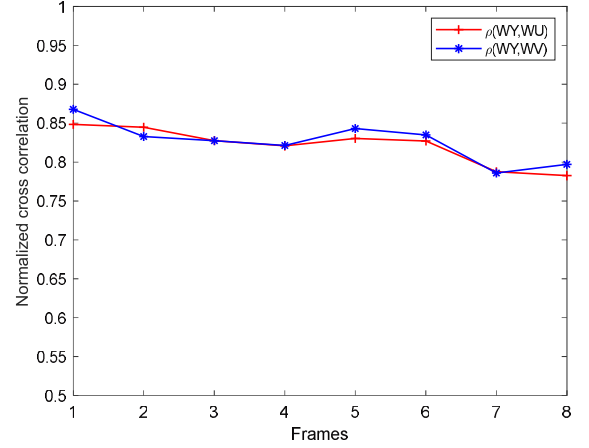


Fig. 2 Normalized cross correlation of 8 frames

determined by the similarity among neighboring luminance pixels, large when $Y(r)$ is similar to $Y(s)$, and small when they are different.

In VVC, due to the partition structure, there are a large amount of blocks with very small size, and most of the pixels in these small blocks are adjacent pixels. Therefore, there may exist strong cross component correlation in a block. To validate this, we select eight frames randomly from video test sequences. For each block in the frames, the W_{rs} on Y and C (either U or V) is obtained by solving following constrained least squares problems:

$$J(Y) = \sum_r [Y(r) - \sum_{s \in N(r)} WY_{rs} \cdot Y(s)]^2 \quad (5)$$

$$J(C) = \sum_r [C(r) - \sum_{s \in N(r)} WC_{rs} \cdot C(s)]^2 \quad (6)$$

where WY_{rs} , WC_{rs} denotes the local correlation coefficients in Y and C . The normalized cross correlation between WY and WC (either WU or WV) is computed and shown in Fig. 2.

From Fig. 2, it can be found that the averaged cross correlation coefficient for 8 frames is greater than 0.8, which indicates that the relation of a pixel and its neighbors in different color channels follows a very similar pattern. Therefore, the U and V components can be predicted using the coloring pattern of the Y component, and this is the main idea of this paper.

However, since Eq. (5) contains too many variables and too few known conditions, it is not feasible to be applied in video coding context. In this paper, a low complexity scheme is designed to utilize the coloring principle in cross component prediction. To reduce the number of pixels in $N(r)$, only some selected position is sampled. In addition, to simply the determination of W_{rs} , a linear mapping method is adopted instead of the least square optimization in Eq. (5).

3.2 Selection of Reference Sample Position

For a block in intra prediction, only the top adjacent pixels and left adjacent pixels of the block are available. To reduce the number of neighboring pixels involved in coloring

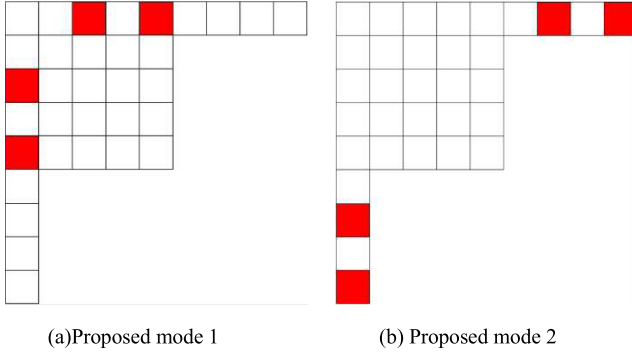


Fig. 3 Sample position of 4x4 block

process, only N positions are sampled. To balance between computation complexity and prediction efficiency, N is set to 4.

The criterion to select the optimal sample positions is to minimize the distance between sampled pixels and the pixels to be predicted. Assuming that the current block size is $W \times H$, the upper adjacent pixel positions are represented by $(0, -1), (1, -1) \dots (W + H - 1, -1)$, and the left adjacent pixel positions are represented by $(-1, 0), (-1, 1) \dots (-1, W + H - 1)$. Two modes of sample positions are designed in this paper. For mode 1, the selected positions are $((W+H)/8, -1), (3(W+H)/8, -1), (-1, (W+H)/8)$, and $(-1, 3(W+H)/8)$, and the selected positions are illustrated in Fig. 3 (a) as an example for 4x4 block. For this mode, the sample positions may benefit pixels located in top-left part of the block. For mode 2, the selected positions are $(5(W+H)/8, -1), (7(W+H)/8, -1), (-1, 5(W+H)/8)$, and $(-1, 7(W+H)/8)$, and these positions are illustrated in Fig. 3 (b) as an example for 4x4 block. For this mode, the sample positions may benefit pixels located in bottom-right part of the block.

3.3 Determination of the Correlation Coefficients

As described in Sect. 3.1, correlation coefficients WY_{rs} are important parameters in colorization process. In [12], WY_{rs} is determined by the similarity existed in luminance pixels. In this paper, the same clue is followed. To reduce the computation complexity, a lightweight scheme based on linear mapping is designed to compute the correlation coefficients.

Intensity difference is used to measure the similarity between the current pixel and the $N(N = 4)$ selected reference pixels. If the difference is small, the similarity is large; otherwise, the similarity is small. Let (i, j) be the position of the pixel to be predicted in chroma components, and define the luminance intensity of the 4 reference sample positions described in Sect. 3.2 as Y_1, Y_2, Y_3, Y_4 . We use the following equations to find the correlation coefficients:

$$YID(i, j) = \sum_{k=1}^4 |Y(i, j) - Y_k| \quad (7)$$

$$WID_1(i, j) = |Y(i, j) - Y_1| / YID(i, j) \quad (8)$$

Table 1 Code for each modes

Mode	VTM12.1	Updated
DM	00	00
LM	10	100
LM_T	110	1010
LM_L	111	1011
DC	0100	0100
Planar	0101	0101
Horizontal	0110	0110
Vertical	0111	0111
Proposed mode 1		110
Proposed mode 2		111

$$WID_2(i, j) = |Y(i, j) - Y_2| / YID(i, j) \quad (9)$$

$$WID_3(i, j) = |Y(i, j) - Y_3| / YID(i, j) \quad (10)$$

$$WID_4(i, j) = |Y(i, j) - Y_4| / YID(i, j) \quad (11)$$

$$WY_1(i, j) = 0.5 - WID_1(i, j) \quad (12)$$

$$WY_2(i, j) = 0.5 - WID_2(i, j) \quad (13)$$

$$WY_3(i, j) = 0.5 - WID_3(i, j) \quad (14)$$

$$WY_4(i, j) = 0.5 - WID_4(i, j) \quad (15)$$

Among the above equations, $YID(i, j)$ represents the total of the absolute intensity difference of Y and is used for normalization. $WID_k(i, j)$ represents the weight of the k -th intensity difference. $WY_k(i, j)$, $k = 1, \dots, 4$ represent the obtained correlation coefficients.

Once $WY_k(i, j)$, $k = 1, \dots, 4$ is obtained, if we define the 4 reference chroma pixels located at the positions described in Sect. 3.2 as C_1, C_2, C_3, C_4 , the chroma pixel located at (i, j) can be predicted using:

$$C(i, j) = \sum_{k=1}^4 WY_k(i, j) \cdot C_k \quad (16)$$

3.4 Code of Modes

The two modes proposed in this paper are used as a supplement to the original chroma prediction modes in VVC. To encode the two modes, the syntax of chroma prediction modes needs to be modified. According to the probability distribution of the chroma prediction modes, the code for each prediction mode is defined as in Table 1.

4. Experimental Results

The proposed modes are implemented on VTM-12.1 [13] test model. The performance of the proposed modes is evaluated based on JVET common test condition (CTC) dataset which includes 20 video clips in 4:2:0 color format. The experiments were done according to the common test condition with all intra (AI) and quantization parameters (QPs) of 22, 27, 32 and 37. The Bjontegaard Delta Rate (BD-Rate) method is used for evaluating the performances. The

Table 2 Bd-rate performance of proposal compared to VTM12.1 anchor

Test Sequence	All Intra (VTM12.1)				
	Y	U	V	EncT	DecT
Class A	-0.02%	-0.54%	-0.71%	102%	100%
Class B	-0.01%	-1.00%	-0.86%	102%	100%
Class C	-0.01%	-0.99%	-1.26%	103%	101%
Class D	0.00%	-1.31%	-1.20%	103%	100%
Class E	0.00%	-0.69%	-0.87%	102%	99%
Overall	-0.01%	-0.92%	-0.96%	103%	100%

BD-Rate results of the proposed method in 4:2:0 is shown in Table 2. As shown in Table 2, the proposed method in All Intra configuration provides 0.92% and 0.96% BD-Rate improvements in U and V components, respectively.

From the complexity perspective, since lightweight scheme is used, the proposed method does not introduce significant encoding runtime to the codec. As shown in Table 2, the encoding runtime overheads are in the range of 2% to 3% in AI, which can be considered as a reasonable trade-off taking into account the additional full RD checks in the encoder. Furthermore, the proposed method does not impose measurable extra decoding runtime to the codec in VTM platform.

5. Conclusion

In this paper, two new chroma prediction modes based on coloring principle is proposed for VVC. Experiment results show that, compared with VTM 12.1, the proposed algorithm has an average of -0.92% and -0.96% BD-rate improvement for U and V components for All Intra (AI) configurations. The complexity increases at both encoder and decoder sides are negligible.

Acknowledgments

This work is supported by Beijing Municipal Natural Science Foundation (No.4202018), Great Wall Scholar Project of Beijing Municipal Education Commission

(CIT&TCD20180304).

References

- [1] Versatile Video Coding, Version 1, Rec. ITU-T H.266, ISO/IEC FDIS 23090-3, July 2020.
- [2] S. De-Luxan-Hernandez, V. George, J. Ma, T. Nguyen, H. Schwarz, D. Marpe, and T. Wiegand, "An intra sub-partition coding mode for VVC," 2019 IEEE International Conference on Image Processing (ICIP), pp.1203–1207, IEEE, 2019.
- [3] 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, IEEE, 2019.
- [4] B. Bross, P. Keydel, H. Schwarz, D. Marpe, and T. Wiegand, "Multiple reference line intra prediction," Document JVET-L0283, Oct. 2018.
- [5] L. Zhao, X. Zhao, S. Liu, X. Li, J. Lainema, G. Rath, F. Urban, and F. Racape, "Wide angular intra prediction for versatile video coding," 2019 Data Compression Conference (DCC), pp.53–62, IEEE, 2019.
- [6] S.H. Lee and N.I. Cho, "Intra prediction method based on the linear relationship between the channels for YUV 4:2:0 intra coding," IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, pp.1037–1040, Nov. 2009.
- [7] K. Zhang, J. Chen, L. Zhang, X. Li, and M. Karczewicz, "Enhanced cross component linear model for chroma intra prediction in video coding," IEEE Trans. Image Process., vol.27, no.8, pp.3983–3997, Aug. 2018.
- [8] L. Zhang, W. Chien, J. Chen, X. Zhao, and M. Karczewicz, "Multiple direct mode for intra coding," IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, Florida, USA, Dec. 2017.
- [9] M.G. Blanch, S. Blasi, A. Smeaton, N.E. O'Connor, and M. Mrak, "Chroma Intra Prediction With Attention-Based CNN Architectures," 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, pp.783–787, 2020.
- [10] M.G. Blanch, S. Blasi, A.F. Smeaton, N.E. O'Connor, and M. Mrak, "Attention-Based Neural Networks for Chroma Intra Prediction in Video Coding," IEEE Journal of Selected Topics in Signal Processing, vol.15, no.2, pp.366–377, Feb. 2021.
- [11] L. Zhu, Y. Zhang, S. Wang, S. Kwong, X. Jin, and Y. Qiao, "Deep Learning-Based Chroma Prediction for Intra Versatile Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol.31, no.8, pp.3168–3181, Aug. 2021.
- [12] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," ACM Transactions on Graphics (TOG), vol.23, no.3, pp.689–694, 2004.
- [13] VTM12.1, https://vcgit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM