A Fast Chroma Intra-Prediction Mode Decision Algorithm Based on Texture Characteristics for VVC

Zhi LIU†(a), Member, Yifan SU†, Shuzhong YANG††(b), and Mengmeng ZHANG†††(c), Nonmembers

SUMMARY Cross-component linear model (CCLM) chromaticity prediction is a new technique introduced in Versatile Video Coding (VVC), which utilizes the reconstructed luminance component to predict the chromaticity parts, and can improve the coding performance. However, it increases the coding complexity. In this paper, how to accelerate the chroma intra-prediction process is studied based on texture characteristics. Firstly, two observations have been found through experimental statistics for the process. One is that the choice of the chroma intra-prediction candidate modes is closely related to the texture complexity of the coding unit (CU), and the other is that whether the direct mode (DM) is selected is closely related to the texture similarity between current chromaticity CU and the corresponding luminance CU. Secondly, a fast chroma intra-prediction mode decision algorithm is proposed based on these observations. A modified metric named sum modulus difference (SMD) is introduced to measure the texture complexity of CU and guide the filtering of the irrelevant candidate modes. Meanwhile, the structural similarity index measurement (SSIM) is adopted to help judging the selection of the DM mode. The experimental results show that compared with the reference model VTM8.0, the proposed algorithm can reduce the coding time by 12.92% on average, and increases the BD-rate of Y, U, and V components by only 0.05%, 0.32%, and 0.29% respectively.

key words: VVC, chroma prediction mode, CCLM, texture complexity

1. Introduction

One of the core ideas of the chromaticity coding technique in Versatile Video Coding (VVC) is to use the correlation among the components to improve the coding performance. Since the luminance component has been coded before coding the chromaticity component, much of the coded information of the luminance component, such as the pixel value, the partition information, and the prediction mode of the luminance component can be used to aid the prediction and coding of the chromaticity component. In chromaticity prediction, VVC introduces a new Cross-component linear model (CCLM) prediction technique that establishes a linear model between luminance and chromaticity component, and uses the linearly transformed data of the reconstructed luminance component to obtain the predicted value of the chromaticity component, and improves the prediction efficiency tangibly. However, the efficiency of chromaticity prediction is affected by the texture characteristics of current coding unit (CU). In one hand, for a CU with complex texture, since the distribution of luminance and chrominance components is not flat, CCLM cannot accurately represent the correlation between the components [1]. On the other hand, PLANAR and DC are the default modes placed in chromaticity prediction candidate modes list. However, for CU with complex texture, the probability to select any one of these two modes as the final mode is pretty low. Therefore, there exists relatively high computational redundancy in chromaticity prediction.

Currently, the studies for chromaticity prediction mainly concentrate on improving the coding performance. They can be classified into two kinds. One is to optimize the selection of the direct mode (DM) used in chromaticity prediction candidate modes list [2]–[6]. The other is to optimize the prediction model used in CCLM to improve prediction efficiency [7]–[9].

In this paper, based on the analysis of the distribution of the optimal mode for each class of video sequence, and a fast chroma intra-prediction mode decision algorithm is proposed based on texture characteristics. As far as we know, there is no relevant study on saving the coding time consumed in chromaticity prediction process.

2. Algorithm Design

2.1 Main Idea

The candidate modes list defined in chroma intra-prediction of VVC includes three parts: DM mode, CCLM modes (LM, LM_L, LM_T), and traditional modes (DC, PLANAR, VER, and HOR). During chroma intra-prediction process, the encoder traverses the candidate list to find the optimal mode, which brings high computation complexity. In our study, we find that the final chroma intra-prediction mode for a CU is closely related with the texture characteristics of the area.

The relation between texture complexity of CU and the optimal mode in chroma prediction is studied for class A~E test sequences [10]. For CU with simple texture, the selection probability of each chroma prediction candidate mode is shown in Fig. 1 a. As can be found from the figure, the selection probability of HOR, VER is extremely low, with
2.2 Measures of Texture Characteristics

To utilize the above observations, the key is to find suitable metrics to measure texture complexity and similarity. Some metrics can be used to describe the texture properties of images, such as grayscale co-occurrence matrices, gradient functions, and sum modulus difference (SMD). Through experimental evaluation, we find that both grayscale co-occurrence matrices and gradient functions introduce excessive computational complexity. SMD can provide a good representation of the texture complexity of the image in grayscale, and has small computation complexity. Therefore, SMD metric is adopted to represent texture complexity in this paper, which is described as follows:

\[
\text{SMD}_{\text{orig}} = \sum_{i=1}^{H} \sum_{j=1}^{W} (|f_{i,j} - f_{i,j-1}| + |f_{i,j} - f_{i-1,j}|) \quad (1)
\]

where \(W\) and \(H\) are the width and height of the CU, respectively, and \(f_{i,j}\) is the gray value of the pixel at \((i, j)\).

In order to improve the texture complexity description performance of SMD in chromaticity prediction, in this paper, we use \(p_{i,j}\) to replace \(f_{i,j}\) in (1), which contains both luminance and chrominance information of the pixel located at \((i, j)\). In addition, to reduce the computation complexity, SMD can be calculated once every two pixels. The modified SMD is defined as follows:

\[
\text{SMD} = \frac{4}{W \times H} \sum_{j=0}^{H/2} \sum_{i=0}^{W/2} \left( |p_{2i,2j} - p_{2i,2j+1}| + |p_{2i,2j} - p_{2i+1,2j}| \right) \quad (2)
\]

SMD is calculated in every CU. To judge the texture flatness of a CU, two thresholds are defined, namely \(T_1\) and \(T_2\). When \(\text{SMD} > T_1\), it is judged that the CU texture is complicated. When \(\text{SMD} < T_2\), it is judged that the CU texture is simple. Otherwise, it is judged that the CU texture is normal and need further processing.

The structural similarity index measurement (SSIM) is adopted as the metric to measure the texture similarity between the chrominance component and corresponding luminance component of a CU. SSIM is defined as:

\[
\text{SSIM}(X, Y) = L(X, Y) \cdot C(X, Y) \cdot S(X, Y)
\]
\[
\sigma = \frac{2}\sqrt{u_x u_y + C_1} \cdot \frac{2\sigma_x \sigma_y + C_2}{(\sigma_x^2 + \sigma_y^2 + C_2)} \left( \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \right)
\]

\[
\sigma = \frac{(2u_x u_y + C_1)(2\sigma_x \sigma_y + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}
\]  

(3)

Where \(X\) and \(Y\) are the two images to be compared. \(u_x\) and \(u_y\) represent the average of the images \(X\) and \(Y\) respectively. \(\sigma_x\) and \(\sigma_y\) represent the standard deviations of the images \(X\) and \(Y\) respectively. \(\sigma_{xy}\) represents the covariance of image \(X\) and \(Y\). \(C_1, C_2,\) and \(C_3\) are constants to prevent the denominator from being 0 and maintain stability. Usually take \(C_1 = 6.5025, C_2 = 58.5225, C_3 = 2/2\). In this paper, to reduce the computation complexity, the calculation of \(u_x, \sigma_x^2\) and \(\sigma_{xy}\) is modified as:

\[
u_x = \frac{4}{H \cdot W} \sum_{j=0}^{H/2} \sum_{i=0}^{W/2} X(2i, 2j)
\]  

(4)

\[
\sigma_x^2 = \frac{4}{H \cdot W} \sum_{j=0}^{H/2} \sum_{i=0}^{W/2} (X(2i, 2j) - u_x)^2
\]  

(5)

\[
\sigma_{xy} = \frac{4}{H \cdot W} \sum_{j=0}^{H/2} \sum_{i=0}^{W/2} (X(2i, 2j) - u_x)(Y(2i, 2j) - u_y)
\]  

(6)

The range of SSIM is \([0, 1]\), and the threshold to judge the similarity is defined as \(T_3\). When \(SSIM \geq T_3\), the texture of the chrominance block and the corresponding luminance block are similar, and the DM mode is directly selected as the optimal mode for chrominance prediction.

### 2.3 Algorithm Description

The proposed algorithm optimizes the chroma prediction mode decision by comparing SMD with two thresholds \(T_1\) and \(T_2\) to judge the texture complexity, and by comparing SSIM with threshold \(T_3\) to judge the texture similarity. When \(SMD > T_1\), it is determined that the CU texture is complex, and four candidate modes LM, LM, T, DC and PLANAR are disabled. In this case, there is no need to further calculate SSIM, and encoder only need to compare the rate-distortion costs of DM, LM, HOR and VER modes to find the optimal one. The algorithm is described as follows:

1. Get the texture complexity value SMD of the current CU.
2. When \(SMD > T_1\), it is determined that the CU texture is complicated, and the four modes namely LM, LM, T, DC, and PLANAR in the chroma intra-prediction candidate mode are disabled, and the rate-distortion costs of the remaining four modes (DM, LM, HOR, VER) are compared.
3. When \(SMD \leq T_1\), get the texture similarity value SSIM value of current CU. If \(SSIM \geq T_3\), determine whether the DM mode is the optimal mode for chroma prediction directly.
4. When \(SMD < T_2\) and \(SSIM < T_3\), it is determined that the CU texture is relatively simple, disable the HOR and VER modes.
5. Otherwise, it is determined that the CU texture complexity is normal, performing the original prediction process.

The algorithm flow chart is shown in Fig. 3:

### 2.4 The Selection of Thresholds

Three thresholds, \(T_1, T_2,\) and \(T_3\) have been defined in the proposed algorithm. Thresholds \(T_1\) and \(T_2\) are used to judge the texture complexity of a CU, threshold \(T_3\) are used to determine whether the DM mode is the optimal mode. To find optimal \(T_1, T_2,\) and \(T_3\), videos from each sequence of different resolutions are selected for experimental training. For thresholds \(T_1\) and \(T_2\), we first fix the value of \(T_1,\) and analyze the impact of different \(T_2\) on the performance of the algorithm, and obtain the optimal threshold \(T_2\). Then, we
change the value of $T_1$ to find its impact on the performance of the algorithm, and obtain the optimal $T_1$ and $T_2$.

For the threshold $T_3$, we compare its influence on the algorithm performance under different values to find the optimal one.

The chromaticity information consists of a U component and a V component, and it has been found in our study that the effect of the threshold on the U and V components is very similar. Due to space limitation, only the impact of V component is illustrated in this paper.

As can be found from Fig. 4, when the value of $T_1$ is 80, the value of $T_2$ is 17, the algorithm has the lowest deterioration of the BD-rate. When the value of $T_3$ is greater than 0.9, the performance of the algorithm becomes stable, so the threshold $T_3$ is set to 0.9.

3. Experimental Results

In order to evaluate the performance of the proposed algorithm, it is implemented in VVC reference software VTM8.0. For each test sequence, five frames are tested in different QPs (22, 27, 32, 37). The encoding configuration is All Intra Main 10. To analyze the performance of the algorithm, the three-component BD-rate increments $\Delta BR_1$ for Y, $\Delta BR_2$ for U, and $\Delta BR_3$ for V of Y, U, and V and the coding time saving percentage $\Delta ET$ are used as the measurement standard.

In the experiments, standard test sequences of HEVC are used, which involve different scenes and different resolutions. Since there is no relevant research on the fast algorithm of chroma-prediction mode, the performance of the VTM8.0 is used as the reference. The experimental results are shown in Table 1. The results show that the coding complexity of chroma-intra-prediction is greatly reduced in the propose algorithm. It achieves a 12.92% time saving on average with the BD rates of Y, U, and V components have only increased by 0.05%, 0.32%, and 0.29% respectively.

### References

[1] 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.

[2] T.-D. Chuang, C.-Y. Chen, et al., “CE1-related: Separate trees for intra slices (without multi-DMs) with an implicit split to 64x64,” JVET-K0230, 11th JVET Meeting: Ljubljana, SI, 10–18 July 2018.

[3] N. Choi, M. Park, and K. Choi, “CE3-related: Chroma DM modification,” JVET-L0053, 12th Meeting: Macao, CN, 3–12 Oct. 2018.

[4] L. Zhang, K. Zhang, et al., “CE3-related: Modified chroma derived mode,” JVET-L0272, 12th Meeting: Macao, CN, 3–12 Oct. 2018.

[5] L. Zhang, W.-J. Chien, J. Chen, et al., “EE5: Multiple direct modes for chroma intra coding,” JVET-E0062, 5th Meeting: Geneva, CH, 12–20 Jan. 2017.

[6] N. Choi, M. Park, and K. Choi, “CE3-related: Chroma DM modification,” JVET-L0053, 12th Meeting: Macao, CN, 3–12 Oct. 2018.

[7] J. Li, M. Wang, L. Zhang, K. Zhang, S. Wang, S. Wang, S. Ma, and W. Gao, “Sub-Sampled Cross-Component Prediction for Chroma Component Coding,” 2020 Data Compression Conference (DCC), Snowbird, UT, USA, pp.203–212, 2020.

[8] X. Ma, H. Yang, J. Chen, “Tests of cross-component linear model in BMS1.0,” JVET-K0190, 11th JVET Meeting: Ljubljana, SI, 10–18 July 2018.

[9] G. Laroche, J. Tiaquet, C. Gisquet, and P. Onno, “CE3: Cross-component linear model simplification (Test 5.1),” JVET-L0191, 12th Meeting: Macao, CN, 3–12 Oct. 2018.

[10] F. Bosson, J. Boyce, K. Suehring, X. Li, and V. Seregin, “JVET common test conditions and software reference configurations for SDR video,” JVET-L0101, 12th Meeting: Macao, CN, 3–12 Oct. 2018.

[11] A. Filippov, V. Rufitskiy, J. Chen, and E. Alshina, “Intra Prediction in the Emerging VVC Video Coding Standard,” 2020 Data Compression Conference (DCC), Snowbird, UT, USA, p.367, 2020.