

# RDO-based Secondary Prediction Scheme for Image Compression

Hongkui Wang<sup>1</sup>, Junhui Liang<sup>1</sup>, Yamei Chen<sup>1</sup>, Hailang Yang<sup>1</sup>, Shengwei Wang<sup>1</sup>, Li Yu<sup>1\*</sup>,  
 Ning Wang<sup>2</sup>, Zhengang Li<sup>2</sup>

<sup>1</sup>School of Electron. Inf. & Commun., Huazhong Univ. of Sci. & Tech., <sup>2</sup>ZTE Corporation  
 hustlyu@hust.edu.cn<sup>1</sup>, {wangning, li.zhengang1}@zte.com.cn<sup>2</sup>

## Abstract

*Image compression plays an important role in information transmission. To further improve its efficiency, this paper proposes an rate-distortion optimization (RDO) based secondary prediction scheme. In this scheme, the block is predicted twice and the optimal prediction is chosen out with a content-based RDO. Specifically, the content property is introduced into prediction and a content-based RDO is proposed at first. Then, a secondary prediction scheme is developed and the block is predicted once again with a distance-based bi-directional prediction method. In order to verify the effectiveness of our scheme, it is incorporated into the latest test model of versatile video coding standard for image compression. Compared with existing image codecs, experimental results show that our codec has the lowest decoding complexity and achieves best quality at the similar compression ratio.*

## 1. Introduction

With the exponentially increasing demand of digital image and video services, an ever stronger demand for image compression has been created, which is aiming at achieving highly compact representation for images and videos by exploiting various types of redundancies [6]. Classical image compression standards usually compress images by reducing three kind of redundancies within images, i.e., spatial redundancy, statistical redundancy, and visual redundancy.

JPEG and JPEG2000 are two image compression standards released by Joint Photographic Experts Group committee. As the common image compression standards, they have been widely applied for image transmission, which save the storage space and transmission bandwidth greatly. In 2012, the BPG codec [2] based on the High Efficiency Video Coding (HEVC) standard [6] was proposed, which achieves better performance than JPEG and JPEG2000. In last year, the newest video coding standard named Ver-

satile Video Coding (VVC) [3] was released. Compared with HEVC, more new coding technologies have been introduced into this standard, which leads to a great improvement in video compression. Compared with BPG, image compression with this standard will further improve the compression ratio effectively.

In image compression, prediction plays an important role in reducing spatial redundancy. Due to the reference samples for image prediction locate at the above and left sides of the block, the correlation among samples generally decrease with the distance increase, which results in the performance of prediction near the bottom-right corner of the block is poorer.

To improve this situation, a novel image prediction scheme called RDO-based secondary prediction (RDO-SP) is proposed in this paper. In this scheme, the block is predicted twice and the optimal prediction is chosen out with a content-based rate-distortion optimization (RDO). The first prediction is the same as the traditional prediction in VVC. In second prediction, the reference samples for this prediction come from the boundary pixels of the current reconstructed block. The distance-based bi-directional prediction method is designed to obtain the second prediction block. To ensure that images with different content have similar quality, an adaptive Lagrange multiplier model is constructed to achieve content based rate-distortion optimized for image compression. Finally, the optimal prediction (i.e., the first prediction or the second prediction) is picked out under the content-based RDO criterion. To verify the effectiveness of the proposed scheme, it has been incorporated into the test model of VVC standard (VTM). RDO-SP scheme will increase the complexity of VTM-based codec. To reduce its complexity, our VTM-based codec is simplified to a pure image codec which has lower complexity.

## 2. Image Prediction in the VTM

In image compression, the predicted image can be obtained by intra prediction. In this section, we will review the partition of the CTUs and the intra prediction in VTM.

\*This work was supported by the NSFC (NO. 61871437)

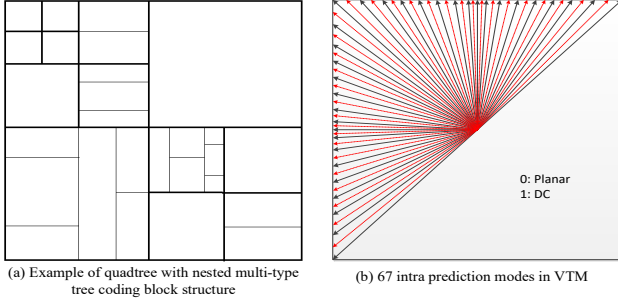


Figure 1. Partition of CTUs and intra prediction modes [3].

### 2.1. Partitioning of the CTUs using a tree structure

In HEVC, the frame of a sequence is divided into coding tree units (CTUs), and CTU can be partitioned into coding units (CUs) via a quad-tree structure. Each CU can be further split into one, two or four prediction units (PUs) according to the PU splitting type. After obtaining the residual block by applying the prediction process based on the PU splitting type, a CU can be partitioned into transform units (TUs) according to another quad-tree structure similar to the coding tree for the CU. One of the key feature of HEVC partition is that it has the multiple partition conceptions including CU, PU, and TU. Unlike HEVC, VVC removes the separation of CU, PU, and TU conception, and supports more flexibility for CU partition shapes. Figure 1-(a) shows a CTU divided into multiple CUs with a quad-tree and nested multi-type tree coding block structure, where the bold block edges represent quad-tree partitioning and the remaining edges represent multi-type tree partitioning. Specific partition rules can be found in [3].

### 2.2. Image Prediction in the VTM

Usually, in intra prediction, the current block is predicted based on the single reference line with the limited prediction modes. The optimal mode is picked out based on the rate-distortion optimization (RDO). However, as for the complex block with multiple directions or with complicated texture, it is difficult to obtain high prediction precision. To improve this situation, compared with HEVC, VVC extends prediction modes from 35 to 67. The new directional prediction modes are shown as red dotted arrows in Figure 1-(b), and the Planar and DC modes remain the same. These denser directional intra prediction modes apply for all block sizes and both luma and chroma intra predictions [3].

In HEVC, reference samples of the current block come from the nearest reconstructed row or column. Yet, in some situation, the non-adjacent reconstructed pixels may obtain better prediction performance. So, multi-line-based intra prediction [5] and line-based intra prediction [1] have been proposed and their simplified algorithms have been incorporated into the latest version of VTM (i.e., VTM4.0). The

simplified algorithms are named Multiple Reference Line (MRL) and Intra Sub-Partitions (ISP). In MRL, three of the adjacent reconstruction lines are chosen as the candidate reference lines in the encoding loop. In ISP, the current block is divided into line-based sub-partitions and the preceding sub-partition is first coded and reconstructed. Then, the reconstructed sub-partition is applied as the reference samples for the later sub-partitions. These new intra prediction methods improve the compression ratio significantly.

## 3. Proposed Algorithm for Image Compression

To further improve the compression efficiency, content property of image texture is introduced into image prediction and our prediction scheme is proposed in this section.

### 3.1. Content-based rate-distortion optimization

According to priori knowledge, mean of pixel and entropy of image can be used to describe the content property of images. Compared with smooth regions, regions contained complex content are difficult to predict and have more distortion in compressed images. To ensure that areas with different contents have similar quality, in this paper, content-based rate-distortion optimization scheme is achieved by adjusting the Lagrange multiplier properly.

Usually, the difference between pixels is calculated by

$$g_{\Delta}(x, y) = g(x, y) - g(x + \Delta x, y + \Delta y). \quad (1)$$

For a  $N \times N$  CTU,  $g(x, y)$  is one of pixel value in the CTU,  $x$  and  $y$  are the coordinate of the this pixel.  $\Delta x$  and  $\Delta y$  are the offset of the position. Assuming  $g_{\Delta}(x, y) \in [\min, \max]$ , the probability of  $m = g_{\Delta}(x, y)$  is expressed as  $p(m) = \Gamma(m = g_{\Delta}(x, y)) / (N \times N)$ .  $\Gamma$  is the count of  $m = g_{\Delta}(x, y)$ . The mean value of the difference of the current CTU can be calculated by

$$\overline{g_{\Delta}} = \sum_{m=\min}^{\max} m \times p(m). \quad (2)$$

The entropy of the current CTU ( $Ent$ ) can be calculated by

$$Ent = - \sum_{m=\min}^{\max} p(m) \times \log_2(p(m)). \quad (3)$$

In addition, the mean value and standard deviation of the current CTU are computed and represented by  $\mu$  and  $\sigma$ . And finally, the offset of Lagrange multiplier ( $\Delta\lambda$ ) is defined as

$$\Delta\lambda = \min\left(\frac{\sigma}{\overline{g_{\Delta}}} + \frac{Ent/\gamma_1 - \sigma/\overline{g_{\Delta}}}{\gamma_1} + \gamma_2, \gamma_3\right). \quad (4)$$

$\gamma_1, \gamma_2$  and  $\gamma_3$  are the model parameters which are set according to the entropy of each CTU. The new Lagrange multiplier ( $\lambda_{new}$ ) for each CTU is described by

$$\lambda_{new} = \lambda + \Delta\lambda, \quad (5)$$

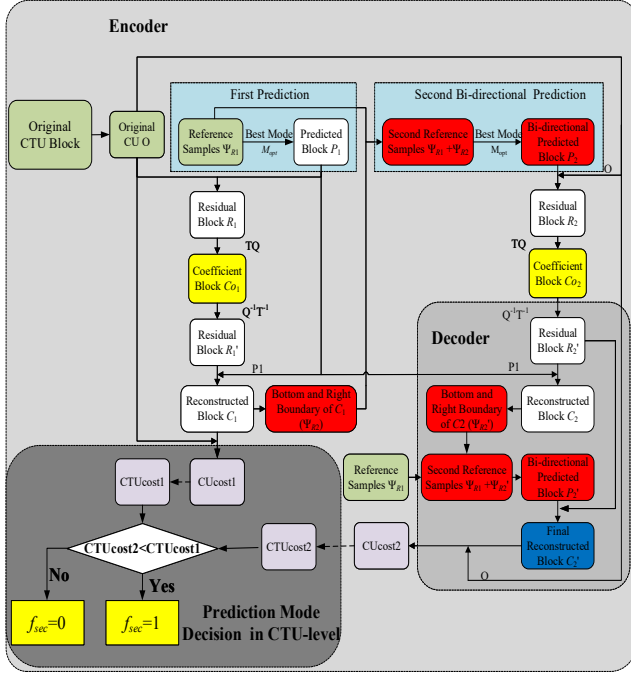


Figure 2. The flowchart of secondary prediction scheme.

here,  $\lambda$  is equal to the default value in VTM. In a same CTU, the RD-cost for each CTU is expressed by

$$J = D + \lambda_{new} \cdot R. \quad (6)$$

$D$  and  $R$  are the distortion and total bits respectively.

### 3.2. RDO-based secondary intra prediction

The reference samples for intra prediction in VTM come from the top and left sides of the current block, which results in the poorer prediction quality in the bottom-right corner. To avoid this limitation, the current block can be predicted once again and the reference pixels for second prediction can be obtained from the reconstructed block of the first prediction. The total flowchart for the proposed secondary prediction is shown in Figure 2.

As shown in Figure 2, the proposed scheme includes three steps (i.e., first prediction, second prediction and the optimal prediction decision). The first prediction is the same as intra prediction in VTM4.0. Based on the top and left reconstructed pixels ( $\Psi_{R1}$ ), all intra modes are traversed and the optimal mode ( $M_{opt}$ ) is selected out with the minimum RD-cost. After transformation, quantization and corresponding opposite operations, the reconstructed block ( $C_1$ ) is obtained. Then, pixels in the bottom and right boundary of  $C_1$  ( $\Psi_{R2}$ ) are picked out and applied as the reference pixels for second prediction combined with  $\Psi_{R1}$ . For a  $W \times H$  block, an example of all reference pixels for second prediction is shown in Figure 3. The mode for second prediction is the same as the first prediction ( $M_{opt}$ ). Af-

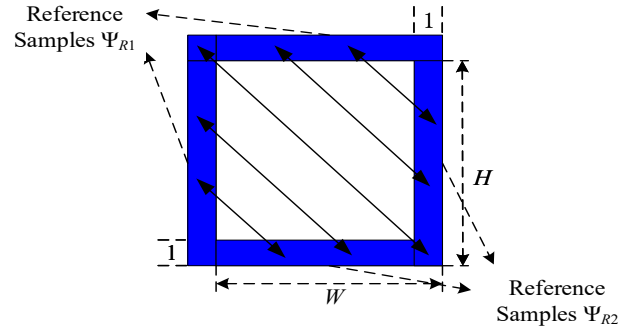


Figure 3. Reference samples for secondary prediction.

ter second prediction, the residual block of second prediction is coded and transferred to decoder. It is worth noting that the bottom and right boundaries of the first predicted block ( $P_1$ ) are the same as that of the second predicted block ( $P_2$ ). Same as encoder, there are two steps of prediction in decoder. Because of the distortion in the transformation and quantization, there is slight difference between the second predicted block in decoder ( $P_2'$ ) and that in encoder ( $P_2$ ).

To maximize compression ratio, content-based rate-distortion optimization proposed in previous section has been incorporated into our secondary prediction. The final prediction (i.e., first prediction or second prediction) is determined with the minimum RD-cost and is indicated by the flag ( $f_{sec}$ ). It should be noting that the proposed secondary prediction scheme is applied in CTU-level for reducing total bits to code this flag.

### 3.3. Distance based bi-directional prediction

As mentioned in Section 2.2, the prediction quality is related on the correlation between the reference pixels and the current pixel. Yet, it is obvious that the correlation generally decrease with the distance increase. Therefore, in proposed secondary prediction scheme, the distance based bi-directional prediction method is designed and applied in second prediction. One example is shown in Figure 4. As shown in this figure, the dark and soft blue pixels mean the original and additional reference samples respectively. Gray pixels, the red pixel and white pixels represent predicted ones, the current pixel and unpredicted pixels respectively. The current pixel's value can be obtained by weighted prediction with reference samples in  $\Psi_{R1}$  and  $\Psi_{R2}$ . For example, if the best intra mode ( $M_{opt}$ ) index is 18, the reference samples for the current pixel are represented by  $\alpha, \alpha_1, \alpha_2$  and  $\beta, \beta_1, \beta_2$ . The predicted value  $g_c$  is calculated by

$$g_c = \frac{d_\alpha(\alpha_1 + 2\alpha + \alpha_2)/4 + d_\beta(\beta_1 + 2\beta + \beta_2)/4}{d_\alpha + d_\beta}, \quad (7)$$

here,  $d_\alpha$  and  $d_\beta$  are the distance between the reference samples and the current predicting pixel.

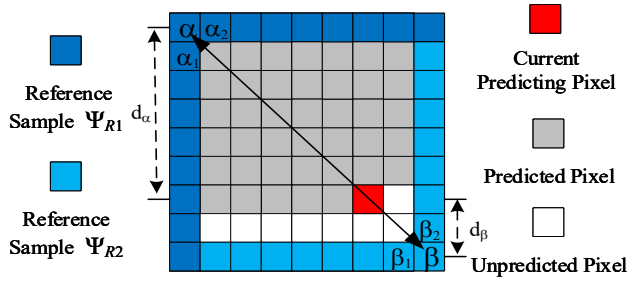


Figure 4. Distance based bi-directional prediction.

## 4. Experimental Results

### 4.1. VTM-based image compression

The proposed RDO-based secondary prediction is incorporated into the latest version of VTM (i.e., VTM4.0). Usually, the image is represented as RGB data with 8-bit sample depth. In order to compress images with VTM4.0, the RGB data is converted to YUV data with the color matrix defined in BT.601 [4]. In encoder and decoder, the YUV data is stored in a 8-bit representation in 4:4:4 chroma format. In addition, after decoding the compressed images, our decoder converts the YUV data back to RGB.

### 4.2. CLIC2019 image compression challenge

We evaluate the performance of the proposed scheme in CLIC image compression challenge, and compare with existing codecs including BPG [2] and original VTM4.0. The performance of three codecs is tested in both the validation dataset and the test dataset [7], including 102 images and 330 images respectively. In low-rate compression track of this challenge, the requirement of this challenge is that the compression is to less than 0.15 bpp across the full dataset. Under this requirement, the performance comparison of three codecs is shown in Table 1 and Table 2. It can be found that our codec (i.e., ZTESmartCodec) achieves best quality with the similar compression ratio. As show in Table 1, compared with the BPG and original VTM4.0, the improvement of PSNR is 1.09 dB and 0.14 dB and the improvement of MS-SSIM is 0.01 and 0.001, respectively. In the test dataset, the improvement of PSNR is 0.99 dB and 0.55 dB and the improvement of MS-SSIM is 0.011 and 0.007. Compared with VTM4.0, it can be found that the improvement of quality in the test dataset is higher than that in the validation dataset. The main reason is that the rate of our codec is closer to requirement of the challenge compared with VTM4.0. In addition, as shown in Table 3, we also calculate the total decoding time for all images in each dataset. Since our decoder is simplified to a pure image decoder, our codec has the least decoding time. In summary, our codec achieves best quality with lowest decoding complexity at the similar compression ratio.

Codecs	Rate(bpp)	PSNR (dB)	MS-SSIM
BPG	0.149	30.84	0.948
VTM4.0	0.149	31.79	0.957
ZTESmartCodec	0.150	31.93	0.958

Table 1. Performance Comparison in the Validation Dataset.

Codecs	Rate(bpp)	PSNR (dB)	MS-SSIM
BPG	0.150	29.60	0.940
VTM4.0	0.134	30.08	0.944
ZTESmartCodec	0.150	30.59	0.951

Table 2. Performance Comparison in the Test Dataset.

Codecs	Validation Dataset	Test Dataset
BPG	80194	260315
VTM4.0	70113	240432
ZTESmartCodec	70079	240315

Table 3. Comparison of Decoding Time in Two Dataset (ms).

## 5. Conclusion

This paper presented an RDO-based secondary prediction scheme for image compression. In this scheme, the block is predicted twice and the best prediction is picked out with the proposed content-based RDO. In second prediction, reference samples come from the boundary pixels of the current reconstructed block. The second prediction block is obtained with the proposed distance-based bi-directional prediction method. This scheme is incorporated into VTM for image compression. Experimental results show that our codec outperforms existing image codecs significantly with the lowest decoding complexity.

## References

- [1] Santiago De-Luxan-Hernandez, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand. Line-based intra prediction for next-generation video coding. In *IEEE International Conference on Image Processing*, 2018.
- [2] F. Bellard. The bpg image format. <http://bellard.org/bpg/>.
- [3] B. Bross, J. Chen, and S. Liu. Versatile Video Coding (Draft 2). Document JVET-K1001 of JVET, Ljubljana. 2018.
- [4] ITU-R. BT.601 studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. <https://www.itu.int/rec/r-rec-bt.601/en>.
- [5] J. Li, B. Li, J. Xu, and R. Xiong. Efficient multiple-line-based intra prediction for HEVC. *IEEE Trans. Circuits Syst. Video Technol.*, 28(4):947–957, 2018.
- [6] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Image Process.*, 22(12):1649–1688, 2012.
- [7] Workshop and challenge on learned image compression (CLIC). <http://www.compression.cc/challenge/>.