LAGRANGIAN METHOD BASED RATE-DISTORTION OPTIMIZATION REVISITED FOR DEPENDENT VIDEO CODING

Xiangwen Wang[*‡], *Li Song*[†‡], *Zhengyi Luo*[*‡], *Rong Xie*[†‡]

[*]College of Electronics and Information Engineering of Shanghai University of Electric Power
[†]Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
[‡]Cooperative Medianet Innovation Center, Shanghai, China
Email: wxw21st@163.com, song_li@sjtu.edu.cn, lzy@shiep.edu.cn, xierong@sjtu.edu.cn

ABSTRACT

Video encoding is based on the DPCM framework where temporal prediction coding introduces Rate-Distortion (RD) dependence. The RD operating point of the current unit depends on the particular choices of RD points of its reference units. Unfortunately, common Lagrangian optimization method based Rate-Distortion Optimization (RDO) for video coding is based on an independence assumption which omits the RD dependences, and thus compromises the RD performance. In this paper, we revisit the Lagrangian optimization method based RDO for dependent video coding. A theoretical RD dependence decoupling method based on independent distortion decomposition is firstly presented. After the discussion of reasonability of the theoretical decoupling method, the practical One Step Ahead Decoupling Strategy (OSADS) is proposed. After implemented on the HEVC encoder, the strategy achieves average 2.1% BD-rate saving compared with the HM encoder under the same low-delay P configuration.

Index Terms— rate distortion optimization, predictive coding, video coding, HEVC

1. INTRODUCTION

Rate-distortion optimization for a video encoder aims to minimize the distortion D subject to a rate constraint R_T. It is described by

$$\min \sum_{l=1}^{L} D_l,\ s.t. \sum_{l=1}^{L} R_l \le R_T \tag{1}$$

where R_l and D_l denote the bit-rate and the distortion of the l-th unit and L is the total number of the coding units. The Lagrangian optimization method is employed to convert the constrained optimization problem to an unconstrained one as

$$\min\{J\},\ J = \sum_{l=1}^{L} D_l + \lambda \sum_{l=1}^{L} R_l, \tag{2}$$

where J and λ are the RD cost and the Lagrangian multiplier, respectively. In practice, usually assuming the rates and the distortions of all the coding units are independent of each other, i.e. under the independence assumption, the full RDO is achieved by individual RDOs as

$$\min\left\{\sum_{l=1}^{L} D_l + \lambda \sum_{l=1}^{L} R_l\right\} = \sum_{l=1}^{L} \min\left(D_l + \lambda R_l\right) \tag{3}$$

Spatial and temporal prediction coding schemes are employed extensively within a video coder. Thus the independence assumption compromises the RD performance inevitably.

Some studies on the dependent RDO have been published. In [1, 2], dynamic programming was introduced for dependent RDO. A temporally dependent RDO scheme was proposed by constructing a temporal distortion propagation chain in [3]. Similarly, in [4] the distortion propagation was transformed to rate influence between dependent CUs under the assumption of high rate quantization. An optimal Lagrangian multiplier calculation method as well as a branch pruning strategy for dynamic programming optimization of predicting dependent rate allocation was proposed in [8, 9].

In this paper we revisit the Lagrangian optimization method based RDO for dependent video coding. The theoretical decoupling method based on independent distortion decomposition is firstly presented along with the discussion of its reasonability. Then, the practical OSADS is proposed. After implemented on the HEVC encoder, the OSADS achieves average 2.1% BD-rate improvement under the same low-delay P configuration.

The rest of the paper is organized as follows. Section 2 reviews the temporal dependence in the video coding framework. Then, the theoretical decoupling strategy based on the assumption of independent distortion decomposition along the distortion propagation chain is presented in Section 3. The new practical decoupling strategy OSADS is proposed in Section 4. Experimental design and some results are given in Section 5. Section 6 draws the conclusions.

This work was supported by Natural Science Foundation of Shanghai (15ZR1417900) and National Natural Science Foundation of China (61671296, 61601282, 61401269 and 61521062).

2. TEMPORAL DEPENDENCE ANALYSIS OF THE VIDEO ENCODER

The video encoder is based on the DPCM framework. Especially, temporal prediction coding is extensively employed. We here focus on the temporal dependence. From the viewpoint of full RDO, the RD operating point of the current CU dependents on the particular choices of RD points of its reference CUs. In other words, the coding options for current CU not only decide its own rate and distortion, but also affect the rates and the distortions of the following CUs which use the current CU as the reference.

It is well known that there is a tradeoff between the rate and the distortion, which implies that the rate and the distortion can be interchanged. Consequently, the RD dependence can be represented by either the distortion propagation or the rate influence along the prediction chain. In this paper, we study the RD dependence from the viewpoint of distortion propagation.

Fig. 1 shows the main procedures of the hybrid video encoder. The T module refers to the transform and the Q module refers to the quantization. I_n is an input Coding Unit (CU) in the nth frame and $\hat{I}_{n-1}$ is the prediction of I_n. Then the residual signal e_n is calculated by $e_n = I_n - \hat{I}_{n-1}$.

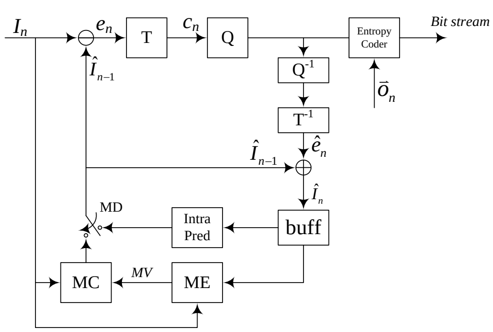

Fig. 1. The hybrid video coding framework

Assuming a certain source distribution and fine quantization of the residual image [5], it is not difficult to deduce that the coding distortion of the current CU can be calculated by

$$\delta_n = e^{-\varepsilon R_n}\left(\sigma^2_{e'_n} + \delta_{n-1}\right) \tag{4}$$

$\sigma^2_{e'_n}$ is the variance of e'_n. $e'_n = I_n - I_{n-1}$ is the prediction residual of the current CU taking the original parent image as the reference. We denote (4) by

$$\delta_n = d_n + \beta_{n\text{-}1}\delta_{n\text{-}1} \tag{5}$$

the first term $d_n = e^{-\varepsilon R_n} \cdot \sigma^2_{e'_n}$ is decided by the coding options of the current CU. So it can be denoted as $d_n \; o_n$. The second term $\beta_{n\text{-}1}\delta_{n\text{-}1} = e^{-\varepsilon R_n} \cdot \delta_{n-1}$ is propagated from the reference frame, which is assumed to be independent of the first term. $\beta_{n\text{-}1}$ is the distortion propagation factor of the (n-1)-th CU to the n-th CU.

We group all CUs depending on each other, i. e., all CUs within one prediction chain, into one dependent CU group. Assuming there are totally N groups, the total rate and distortion can be calculated by

$$D = \sum_{n=1}^{N} D_n, \quad R = \sum_{n=1}^{N} R_n \quad \text{and} \quad D_n = \sum_{m=1}^{M_n} \delta_{nm}, \quad R_n = \sum_{m=1}^{M_n} r_{nm}$$

where M_n is the number of CUs within the n-th dependent CU group. Since there is no dependence between any two groups, the unconstrained RD optimization equals to minimizing N groups separately.

3. THEORETICAL DECOUPLING STRATEGY

We denote the CU number of one dependent CU group by M. Then the optimization for the group can be expressed by

$$\left\{o_1^*, \ldots, o_m^*, \ldots, o_M^*\right\} = \arg\min_{\{o_1,\ldots,o_M\}} \left(\sum_{m=1}^{M} \delta_m + \lambda \sum_{m=1}^{M} r_m \right) \tag{6}$$

o_m^* represents the optimal freedom parameter set for the m-th CU. The freedom parameters include Coding Modes, Reference Index, Motion Vectors and Quantization Index. The transformed coefficient before quantization is also included when SDQ is employed for quantization [6]. Besides these important factors, the Loop Filter parameters and SAO parameters are also critical freedom parameters for the rate and distortion for the state-of-the-art HEVC standard [7].

If all terms in (6) can be decomposed into independent ones, one can decouple the joint optimization into an independent form. In the following, we study the rate term and the distortion term respectively.

We assume any dependent CU can be coded with a fixed rate by selecting the optimal freedom parameters despite the possible distortion change of its reference CUs. This assumption is based the fact that the freedom parameter set of one CU is large enough for an HEVC coder. So each rate term in (6) can be decoupled from its reference CUs and is decided only by the current coding freedom parameters. It can be expressed as

$$r_m \triangleq r_m(o_m) \tag{7}$$

We proceed to study the distortion terms in (6). By employing (5) we can expand the total distortion of the dependent CU group as

$$\sum_{m=1}^{M} \delta_m = \delta_1(o_1) + \sum_{m=2}^{M}\left[d_m(o_m) + \beta_{m-1}\delta_{m\text{-}1}\right] \tag{8}$$
$$= \delta_1(o_1) + \beta_1\delta_1 + d_2(o_2) + \beta_2\delta_2 + d_3(o_3) + \ldots + \beta_{M-1}\delta_{M\text{-}1} + d_M(o_M)$$

For the first CU, the distortion δ_1 is only decided by its coding options o_1 since the first CU is intra coded. This implies $\delta_1 = \delta_1(o_1) = d_1(o_1)$. Assuming the independent de-composition of the distortion (5) can be conducted along the distortion propagation chain recursively, we can derive a general distortion term for the m-th CU by

$$\delta_m = d_m(o_m) + \beta_{m-1}\delta_{m-1} \\ = d_m(o_m) + \sum_{i=1}^{m-1}\prod_{j=i}^{m-1}\beta_j d_i(o_i) \quad (9)$$

Consequently, we may formulate the total distortion of the dependent group by

$$\sum_{m=1}^{M}\delta_m = \delta_1(o_1) + \sum_{m=2}^{M}\left[d_m(o_m) + \beta_{m-1}\delta_{m-1}\right]$$
$$= \left(1+\sum_{j=1}^{M-1}\prod_{i=1}^{M-1}\beta_i\right)d_1(o_1) + \left(1+\sum_{j=2}^{M-1}\prod_{i=2}^{M-1}\beta_i\right)d_2(o_2) + \dots$$
$$+\left(1+\sum_{j=m}^{M-1}\prod_{i=m}^{M-1}\beta_i\right)d_m(o_m) + \dots + (1+\beta_{M-1})d_{M-1}(o_{M-1}) + d_M(o_M)$$
$$= \sum_{m=1}^{M}\left(1+\sum_{j=m}^{M-1}\prod_{i=m}^{M-1}\beta_i\right)d_m(o_m) \quad (10)$$

Now each term within the total distortion is an independent term. Finally, by combining (7) and (10), we can decouple the dependent optimization into an independent form as

$$\min_{o_1,\dots,o_M}\left(\sum_{m=1}^{M}\delta_m + \lambda\sum_{m=1}^{M}r_m\right) \quad (11)$$
$$= \min_{o_1,\dots,o_M}\left(\sum_{m=1}^{M}\left(1+\sum_{j=m}^{M-1}\prod_{i=m}^{M-1}\beta_i\right)d_m(o_m) + \lambda\sum_{m=1}^{M}r_m(o_m)\right)$$
$$= \sum_{m=1}^{M}\min_{o_m}\left\{\left(1+\sum_{j=m}^{M-1}\prod_{i=m}^{M-1}\beta_i\right)d_m(o_m) + \lambda r_m(o_m)\right\}$$

note that the distortion term in (11) is d_m but not δ_m, which is different from Li's strategy [4]. Here δ_m contains the distortion of current CU d_m as well as the distortion propagated from its reference CUs.

The RDO dependence decoupling solution in (11) is based on the assumption that the distortion independence decomposition (5) can be conducted recursively along the distortion propagation chain. Unfortunately, this assumption may be neither reasonable nor practical. Firstly, the calculation of the quantization distortion of one residual image with noise induced by the reference image is still an open problem. Furthermore, as coding options like SDQ, LP and SAO are adopted in the new standard HEVC [7], calculating the propagation factor of one CU to its following CUs cannot be accurate. Secondly, in the RDO procedure for the current unit, the distortion influence on all of its following CUs is prone to be over optimized in the sense of full RDO. Lastly, the whole prediction chain, which usually involves all frames, has to be determined in advance even for the very first CU RDO. Therefore we have to take further efforts for practical RDO decoupling.

4. ONE STEP AHEAD DECOUPLING STRATEGY

From the viewpoint of full RDO, if we have considered the distortion propagation factor between the m-th CU and the (m+1)-th CU when coding the m-th CU, we don't need to consider this factor once more when coding the (m+1)-th CU, since the influence has already been taken into account when coding the m-th CU. This means we only need to consider the dependence in one direction. Based on this observation, we revisit the joint RDO for two dependent CUs.

The joint RDO for the m-th CU and its corresponding (m+1)-th CU which uses the m-th CU as the reference can be formulated by

$$\min_{o_m,o_{m+1}}\left(J_m(o_m) + J_{m+1}(o_{m+1},o_m)\right) \quad (12)$$

Note that the m-th CU is not necessarily an Intra CU. The reason for denoting its RD cost by $J_m(o_m)$, which means the RD cost J_m is only decided by o_m, is that the influence from previously referenced CUs has already been considered.

We re-write the joint optimization as

$$\min_{o_m,o_{m+1}}\left(J_m(o_m) + J_{m+1}(o_{m+1},o_m)\right) \\ = \min_{o_m}\left(J_m(o_m) + J_{m+1}(o^*_{m+1},o_m)\right) \quad (13)$$

where $J_{m+1}(o^*_{m+1},o_m)$ is the minimum Lagrangian cost of the dependent CU when the referenced CU is coded with o_m. It is calculated by

$$J_{m+1}(o^*_{m+1},o_m) = \delta_{m+1}(o^*_{m+1},o_m) + \lambda r_{m+1}(o^*_{m+1},o_m) \quad (14)$$

where $\delta_{m+1}(o^*_{m+1},o_m)$ is the distortion of the dependent CU with the optimal coding option o^*_{m+1} assuming the referenced CU is coded with o_m. By employing the distortion independence decomposition as (5), we can decompose the distortion of the (m+1)-th CU as follows

$$\delta_{m+1}(o^*_{m+1},o_m) = \delta_{m+1}(o^*_{m+1},\delta_m) = \beta_m\delta_m(o_m) + d_{m+1}(o^*_{m+1}) \quad (15)$$

The first equation in (15) is based on the fact that the influence from the coding options of the reference CU can be represented by its distortion term. By substituting (14) and (15) into (13) and employing (7), we obtain

$$\min_{o_m,o_{m+1}}\left(J_m(o_m) + J_{m+1}(o_{m+1},o_m)\right) \quad (16)$$
$$= \min_{o_m}\left(\delta_m(o_m) + \lambda r_m(o_m) + \delta_{m+1}(o^*_{m+1},o_m) + \lambda r_{m+1}(o^*_{m+1},o_m)\right)$$
$$= \min_{o_m}\left(\delta_m(o_m) + \lambda r_m(o_m) + \beta_m\delta_m(o_m) + d_{m+1}(o^*_{m+1}) + \lambda r_{m+1}(o^*_{m+1})\right)$$

Obviously, the term $d_{m+1}(o^*_{m+1}) + \lambda r_{m+1}(o^*_{m+1})$ is dependent only on o^*_{m+1} and independent of o_m. Ultimately, we can decouple the joint dependent optimization as follows

$$o^*_m = \arg\min_{o_m} J_m(o_m) = \arg\min_{o_m}\left(\delta_m(o_m) + \lambda r_m(o_m) + \beta_m\delta_m(o_m)\right)$$
$$= \arg\min_{o_m}\left((1+\beta_m)\delta_m(o_m) + \lambda r_m(o_m)\right) \quad (17)$$
$$= \arg\min_{o_m}\left(\delta_m(o_m) + \frac{\lambda}{(1+\beta_m)}r_m(o_m)\right)$$

The above deduction shows how the joint optimization of two dependent CUs is decoupled. In this decoupling strategy, the accumulated influence on one CU from all the coding options of its parent CUs along the prediction chain is represented by the distortion influence from the direct reference CUs. The same decoupling strategy can be applied to cases where more than two CUs are considered. As shown in (17), simply by adjusting the Lagrangian multiplier, the original Lagrangian optimization form for independent sources is now suitable for dependent video coding.

5. EXPERIMENTAL DESIGN AND SOME RESULTS

5.1. Experimental design

The key to the application of the decoupling strategy is to calculate the propagation factor. The distortion propagation factor is closely related to the referenced times of the CU, the variance of the residual and the QP value of the future referencing CU. In this experiment, forward motion estimation with fixed block size (16x16 in the paper) based on the original frames is conducted to estimate the propagation factor.

Two factors—the referenced times and the weight factor are counted to calculate the propagation factors. The total referenced times, denoted by *T*, refers to the times a CU is referenced. Since the referenced block may stride across several CUs while each coding CU is CU size aligned, the referenced times is counted for every pixel. When coding one CU, the referenced times of all the pixels within the CU is summed up. The weight factor $k_{(y,x,t)}$ takes into account the coding coefficients of the future *t*-th referencing CU. We classify coding CUs into three categories according to the variance. They are the all-zero CUs, the CUs with non-zero coefficients and small residual image variance and the CUs with non-zero coefficients and large residual image variance. The first category corresponds to the SKIP mode and the all-zero mode. We adopt the scheme in [10] to predict the all-zero CUs based on the variance of the residual image. $k(y, x, t)$ is 0.5, 0.25 and 0.125 for the three categories, respectively. The ultimate factor is calculated by

$$\beta_m = \frac{w * \sum_{y=1}^{Y} \sum_{x=1}^{X} \left(\sum_{t=1}^{T} k_{(y,x,t)} \right)}{Y * X},$$

where *Y* and *X* are the vertical and horizontal sizes of the coding CU, respectively. *w* is another weighting factor equal to 0.8 in this paper.

In addition, we employ a QP adaption scheme to adjust the QP value according to the propagation factor as follows

$$QP_m = \begin{cases} QP_g + 2, & if\ \beta_m < 0.2 \\ QP_g, & \\ QP_g - 2, & if\ \beta_m > 0.5 \end{cases},$$

where QP_g refers to the pre-set QP value for the sequence.

5.2. Preliminary results

We implement OSADS in the HEVC reference software HM 16.10 under the Low delay P configuration. To verify the performance of the strategy, we disable the Intra coding modes for the inter frames. The reference frame structure in the motion estimation procedure for calculating the propagation factor is the same as the default configuration, except that only the nearest two reference frames are allowed. The reference frame is decided according to the RD cost, which is the sum of the SAD and the logarithm of the difference motion vectors (DMV) and the reference index multiplied by the Lagrangian multiplier.

The first 100 frames of five Class B sequences are selected for evaluation with the default configuration except that the Intra Coding mode is disabled for Inter frames acting as the anchor. Table I tabulates the evaluation results. We can see that the proposed OSADS achieves up to 2.8% BD-rate saving and 2.1% BD-rate saving on average than the anchor encoder.

TABLE I. RD performance improvement by OSADS in terms of BD-Rate saving.

Sequence	BD-rate saving
Kimono	−1.8%
ParkScene	−2.8%
Cactus	−2.4%
BasketballDrive	−2.0%
BQTerrace	−1.7%
Average	−2.1%

6. CONCLUSIONS

This paper revisits the Lagrangian method based RDO for dependent video coding. The theoretical decoupling method based on independent distortion decomposition is firstly derived. Its reasonability and its practicability are discussed. The simple yet effective OSADS is proposed as a practical solution of full RDO. It can achieve average 2.1% BD-rate improvement under the same low-delay P configuration for the HEVC encoder.

7. REFERENCES

[1] A. Ortega, K. Ramchandran, "Rate-distortion methods for image and video compression", *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23-50, Nov. 1998.

[2] K. Ramchandran, A. Ortega, M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders", *IEEE Trans. on Image Proc.*, vol. 3, pp. 533-545, Sept. 1994.

[3] T. Yang, C. Zhu, X. Fan, Q. Peng, "Source distortion temporal propagation model for motion compensated video coding optimization", *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, pp. 85-90, Jul. 2012.

[4] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, M. T. Sun, "Lagrangian Multiplier Adaptation for Rate-Distortion Optimization With Inter-Frame Dependency", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 117-129, Jan. 2016.

[5] N. S. Jayant, and P. Noll, Digital coding of waveforms, *Englewood Cliffs, NJ: Prentice-Hall*, 1984.

[6] E.-H. Yang, X. Yu, "Rate distortion optimization for H.264 interframe coding: A general framework and algorithms", *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1774-1784, Jul. 2007.

[7] G. J. Sullivan, J. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[8] G. Cheung et al., "On Dependent Bit Allocation for Multiview Image Coding with Depth-Image-based Rendering," *IEEE Transactions on Image Processing*, vol.20, no.11, November 2011.

[9] A. De Abreu et al., "Optimal Lagrange Multipliers for Dependent Rate Allocation in Video Coding," *arXiv*, March 2016, https://arxiv.org/abs/1603.06123.

[10] X. Zhou, Z. H. Yu, and S. Y. Yu, “Method for detecting all-zero DCT coefficients ahead of discrete cosine transform and quantization”, *Electron. Lett.*, vol, 34, no. 19, pp. 1839-1840, Sep. 1998.