# Lagrange Multiplier Based Perceptual Optimization for High Efficiency Video Coding

Shanshe Wang[*], Siwei Ma[†], Debin Zhao[*] and Wen Gao[†]
[*] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China.
E-mail: sswang@jdl.ac.cn; dbzhao@hit.edu.cn Tel: +86-10-62753424
[†] Institute of Digital Media, Peking University, Beijing, China.
E-mail: swma@pku.edu.cn; wgao@pku.edu.cn Tel: +86-10-62753424

***Abstract*—In this paper, a Lagrange multiplier based perceptual optimization scheme is proposed for high efficiency video coding (HEVC). Firstly, we propose a non-linear adjustment scheme for the Lagrange multiplier based on the specialized coding structure of HEVC and human visual system (HVS) characteristic. Then the Lagrange multiplier is incorporated into the rate distortion optimization for mode decision. Experimental results demonstrate the proposed scheme can significantly improve the subjective video quality. Compared to the HM anchor, the maximum subjective quality improvement can be up to 30% and the average subjective performance improvement is over 7%, 11% and 10% for random access, low delay P and low delay B testing configurations respectively.**

## I. Introduction

High Efficiency Video Coding (HEVC) [1] significantly improves the coding efficiency over the preceding standards, such as H.264/AVC [2]. With the MOS as the subjective measurement, the subjective performance improvement can be up to 57% compared to H.264/AVC [3]. For the contribution to the significant coding efficiency improvement of HEVC, rate distortion optimization (RDO) plays an important role besides the new coding tools, e.g. quad-tree partition based coding structure *etc*. The RDO scheme can be described as minimizing the perceived distortion $D$ with the number of used bits $R$ subjected to a constraint $R_c$ as follows.

$$\min\{D\} \qquad s.t. R \le R_c , \tag{1}$$

where $D$ is in terms of the sum of absolute difference (SAD) or the mean square error (MSE).

The above constrained optimization problem can be equally converted to an unconstrained optimization problem as (2) by Lagrangian optimization technique [4].

$$\min\{J\} \qquad J = D + \lambda \cdot R , \tag{2}$$

where $J$ is called rate-distortion (R-D) cost. $\lambda$ is known as Lagrange multiplier which controls the tradeoff between $R$ and $D$.

In (2), $\lambda$ is commonly considered to have great influence for the coding performance. Thus, in order to achieve optimal R-D performance, it is important to choose reasonable $\lambda$. In the H.264/AVC, the $\lambda$ is suggested to be as follows [16]:

$$\lambda = 0.85 \cdot 2^{\frac{QP-12}{3}} , \tag{3}$$

where QP is the quantization parameter (QP). This suggestion was proposed based on empirical results and typical R-D models [14][15]. It also suggests that $\lambda$ is a function of QP only and therefore is independent of the frame properties, which simplifies the problem but may not result in optimal $\lambda$. In HEVC, $\lambda$ is further adjusted according to the frame depth, and significant coding performance can be achieved by this adjustment. In [9], it is stated that the reasonable adjustment for $\lambda$ can achieve perceptual performance improvement.

In the literature, significant progress has been made to adapt $\lambda$ on frame level when MSE is used as the distortion measure. In [17], Chen et al. developed an adaptive λ estimation algorithm by modeling the $R$ and $D$ in $\rho$ domain, where $\rho$ is defined as the percentage of zero coefficients among quantized transform residuals. In [18], Laplace distribution-based rate and distortion models were established to derive $\lambda$ for each frame dynamically.

However, SAD and MSE are widely criticized for not correlating well with perceived quality since the ultimate receiver of the video is human eye. Therefore, the perceptual based video coding is always a hot topic, and vision model has been considered in the early coding standards [5]. However, human visual system is so complicated that it cannot be represented accurately by a simple model. In the actual video coding, a practicable scheme is to incorporate the perceptual distortion into the R-D cost for mode decision. In the previous researches, many works focused on the formulation of perceptual distortion. In [6], Structure Similarity (SSIM) is proposed to be an effective subjective video quality measurement. Based on SSIM, in [7][8], the perceptual distortion is described as (1-SSIM), and then perceptual RDO can be achieved. In [9], the authors proposed to utilize 1/SSIM as the perceptual distortion. And a perceptual optimization scheme is proposed based on the $\lambda$ for HEVC.

In the recent researches, the divisive normalization theory [7][10] is investigated in the perceptual optimization. It is stated that it can reflect the visual characteristic of human eye at a certain extent [11]. In the image/video processing, the divisive normalization is widely utilized [12][13]. However, the calculation scheme for divisive normalized factor (DNF) is somewhat different. e.g. the local characteristic based scheme [13] and distribution model based scheme [11].

In this paper, a Lagrange multiplier based perceptual optimization scheme is proposed to improve the subjective performance of HEVC. Firstly, considering the perceptual characteristic of HVS, a non-linear zooming scheme is proposed to adjust the divisive normalization factor, and then the adjusted factor is utilized to adjust quantization parameter. Secondly, the adjusted factor is applied to adjust Lagrange multiplier for mode decision in order to improve the subjective performance.

The rest of the paper is organized as follows. Section II shows the proposed Lagrange multiplier based perceptual optimization scheme for HEVC. The experimental results are provided in Section III. Finally, we conclude the paper in the last section.

## II. Proposed Perceptual Rate-Distortion Optimization for HEVC

In [7], based on the divisive normalization theory, a perceptual rate-distortion optimization scheme is proposed for H.264/AVC. The main idea is to calculate a divisive normalization factor to adjust the QP for a given macro block. Since HEVC adopts many new coding techniques, it is not reasonable to utilize this method into HEVC directly. Thus in [7], the author improved the scheme in [7] and subjective performance improvement is achieved. However, the improved scheme still does not fully consider the perceptual characteristic of HVS. In this paper, we further improve the perceptual optimization scheme for HEVC based on Lagrange Multiplier.

### *A. Proposed non-linear zooming scheme*

In [7], in order to smooth the DNF, Gaussian filter is firstly to filter the coding unit. Then DNF is calculated by (4).

$$f' = \frac{1}{l}\sum_{i=1}^{l}\sqrt{\frac{\sum_{k=1}^{N-1}(2\cdot I_i^g(k)^2)}{N-1}+C} \Big/ E\left(\sqrt{\frac{\sum_{k=1}^{N-1}(2\cdot I_i^g(k)^2)}{N-1}+C}\right) \quad ,(4)$$

where $I_i^g(k)$ denotes the Gaussian filtered coefficients. $E$ denotes the mathematic expectation. Then the achieved DNF is incorporated into the adjustment of the quantization step, $Q_s$, as follows.

$$Q_s' = f' \cdot Q_s \ , \quad (5)$$

where $Q_s'$ denotes the adjusted quantization step.

However, the DNF after Gaussian filter still fluctuates much as illustrate in Fig. 1. It can be seen that the DNF of adjacent CU varies much and the quantization step of the adjacent CU has a big difference, which lead to the big quality fluctuation of adjacent CU. This evident fluctuation is not accordance to the characteristic of HVS. It can lead to aircraft on the boundary of adjacent CU, which will affect the subjective quality of the frame much.

In order to smooth the fluctuation, we proposed a non-linear zooming scheme to smooth the DNF of adjacent CU as (6). This method can convert the domain with fierce changing into flat domain, which will not bring discomfort for HVS. Meanwhile, the subjective performance can be ensured.

$$f = \frac{3^{f'+1}-1}{2\cdot 3^{f'}+1}, \quad (6)$$

After performing (6), the value of $f'$ can be limited into a suitable domain as [0.5, 1.5]. Fig.2 presents the comparisons between the original DNF distribution and the adjusted DNF distribution by the proposed non-linear zooming scheme. It can be evidently observed that the proposed scheme can obtain much smoother DNF.

Then $f$ is utilized to adjust the quantization step as (6) and the new quantization step will be incorporated to the following quantization process.

$$Q_s' = Q_s \cdot f \ . \quad (7)$$

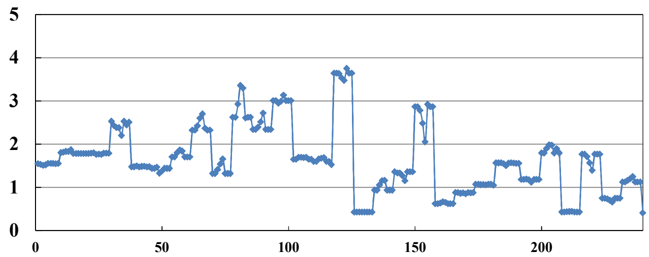


Fig. 1 Distribution of divisive normalization factor in reference [10]

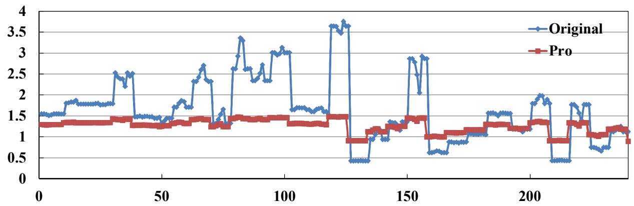


Fig. 2 Distribution comparisons of divisive normalization factor between the proposed non-linear zooming scheme and the scheme in reference [10]

### *B. Lagrange multiplier based perceptual optimization for HEVC*

The shortcoming of traditional rate distortion optimization as in (1) is the component $D$ cannot reflect the characteristic of HVS. The ideal perceptual rate distortion optimization should be represented as:

$$J = D_p + \lambda \cdot R$$

where $D_p$ denotes the perceptual distortion. Intuitively, $D_p$ has great correlation with $D$ and $f$.

For a given coding unit, the distortion can be calculated by

$$\begin{aligned} D_p &= \sum_{i=1}^{l}\sum_{k=0}^{N-1}(C_i(k)'-R_i(k)')^2 \\ &= \sum_{i=1}^{l}\left(\frac{\sum_{k=0}^{N-1}(X_i(k)-Y_i(k))^2}{f^2}\right), \\ &= \frac{D}{f^2} \end{aligned} \quad (8)$$

where $C_i(k)'$ and $R_i(k)'$ denote the divisive normalized value and divisive normalized value of transformed coefficients.

$X_i(k)$ and $Y_i(k)$ are original transformed coefficient and inverse transformed coefficients.

Thus, the perceptual based rate distortion cost can be represented as follow.

$$J = \frac{D}{f^2} + \lambda \cdot R\,, \tag{9}$$

Considering that different adjustment of $\lambda$ for frame in different depth is adopted. Firstly we change the above formula (9) into an equivalent formula as follow.

$$f^2 \cdot J = D + (\lambda \cdot f^2) \cdot R\,. \tag{10}$$

For a given district, the value $f$ is the same. Then in our paper, the defined perceptual rate distortion cost can be modified as:

$$\begin{aligned} J_p &= D + (\lambda \cdot f^2) \cdot R \\ &= D + \lambda_p \cdot R \end{aligned}, \tag{11}$$

where $J_P$ denotes the modified perceptual rate distortion cost. Thus it can be concluded that perceptual rate distortion cost can be considered as an adjustment of Lagrange Multiplier based on perceptual characteristic.

## III. Experimental Results

To evaluate the performance of the proposed algorithm, the experiments are conducted on HEVC test model HM10.0. SSIM is utilized to measure the subjective quality. When comparing the coding performance difference, we utilize the popular method proposed in [19] to calculate the difference between two R-D curves.

### A. R-D performance

Firstly, experiments are conducted for RA configuration. Table I presents the R-D performance of the proposed scheme and the comparisons with related algorithms. It can be seen that the proposed scheme achieved better performance. For class *D* test sequences the subjective coding gain can be up to 10.57%. And the average performance gain is much higher than the other two algorithms. Table II shows the coding performance of the proposed scheme for LDP and LDB configurations. It can be seen that the average coding gains for LDP and LDB are over 11% and 10% on average, respectively. Especially, for the sequence, *BQSquare*, the subjective performance improvement can be over 30% in LDP and LDB testing configuration.

In order to further show the coding efficiency of the proposed scheme, six R-D curves are shown in Fig. 3 for different coding configurations. Compared with HM anchor, it can be observed that the proposed scheme can improve the subjective quality under low and high bit rate.

### B. Subjective quality comparisons

Fig. 4 provides the subjective comparisons between the HM10.0 anchor and the proposed scheme for *BQSquare*. It can be seen that the proposed scheme has better subject quality but with lower bit rate, especially in the areas with more subjective sensitivity.

### C. Complexity analysis

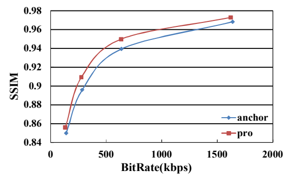
In the proposed perceptual rate-distortion optimization scheme, the calculation of DNF has little influence for the coding complexity. The computation overhead of the proposed method for RA, LDP and LDB is only about 1.34%, 1.2% and 1.17% respctively, which can be neglected.

Table I Subjective performance of proposed scheme and comparisons between other algorithms for RA-main

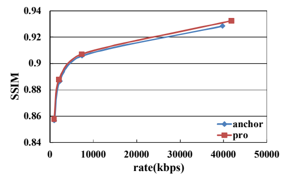
| Sequence | Proposed | | ref [10] | | ref[9] |
|---|---|---|---|---|---|
| | BD -Rate | BD -SSIM | BD -Rate | BD -SSIM | BD -SSIM |
| Class A | 2.19% | -7.67% | 6.46% | -6.19% | -7.7% |
| Class B | 4.33% | -6.74% | 10.15% | -4.54% | -3.6% |
| Class C | 2.36% | -7.01% | 7.93% | -4.33% | -6.8% |
| Class D | 3.04% | -10.37% | 9.15% | -8.90% | -7.3% |
| Class E | 1.64% | -3.77% | 4.36% | -1.74% | -- |
| Arg. | 2.92% | -7.21% | 7.90% | -5.30% | -5.5% |

Table II Subjective performance of proposed scheme for LDP-main and LDB-main

| Sequence | LDP-main | | LDB-main | |
|---|---|---|---|---|
| | BD-Rate | BD-SSIM | BD-Rate | BD-SSIM |
| Class A | 0.79% | -12.72% | 0.43% | -11.99% |
| Class B | 4.78% | -10.43% | 4.17% | -10.11% |
| Class C | 2.21% | -11.09% | 2.13% | -10.48% |
| Class D | 2.70% | -16.27% | 2.48% | -15.38% |
| Class E | 3.40% | -6.43% | 3.51% | -5.74% |
| Arg. | 3.08% | -11.46% | 2.82% | -10.84% |

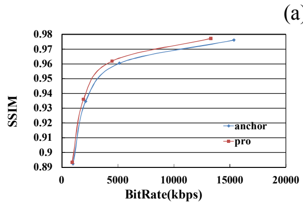



(a) BQSquare_416x240

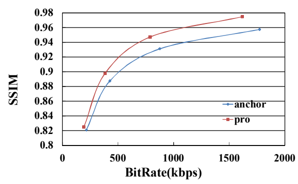



(b) BQTerrance_1920x1080

(a) RA-main

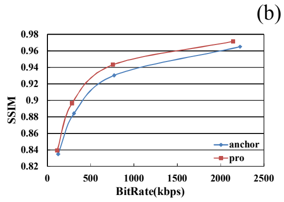



(a) Traffic_2560x1920

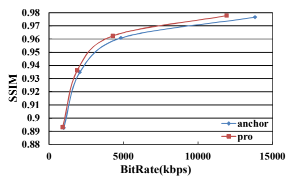



(b) BasketballPass_416x240

(b) LDP-main




(a) BQSquare_416x240




(b) Traffic_2560x1920

(c) LDB-main

Fig. 3 R-D curves of the performance comparison with HM anchor

(a) HM10.0, bitrate：284.7456bit/s，PSNR: 31.8867dB，SSIM: 0.8959

(b) Pro, bitrate：277.2024kbit/s，PSNR: 31.4131dB，SSIM: 0.9094

Fig. 4 Subjective comparisons for *BQSquare*_416x240

## IV. CONCLUSIONS

In this paper, a Lagrange multiplier based perceptual optimization scheme is proposed to improve the subjective quality for HEVC. Firstly, a non-linear zooming scheme is proposed for the calculation of divisive normalization factor. Then the factor is incorporated into the adjustment of the Lagrange multiplier in order to compute the perceptual rate distortion cost for mode decision. The experimental results demonstrate the proposed scheme improved the subjective quality significantly.

## V. ACKNOWLEDGEMENT

This work was supported in part by the National High-tech R&D Program of China (863 Program, 2012AA010805), National Science Foundation of China(61322106, 61103088), which are gratefully acknowledged.

## REFERENCES


[1] G. J. Sullivan, J.-R. Ohm, W. J. Han and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, *IEEE Trans. Circuits Syst. Video Technol.*, vol.22, no.12, pp. 1649-1668, 2012.

[2] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification, ITU-T Rec. H.264 and ISO/IEC 14 496-10 AVC, Joint Video Team, Mar. 2003.

[3] “JCT-VC AHG report: HM subjective quality investigation,” JCT-VC H0022, Feb, 2012.

[4] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[5] I. DalgNc and F. A. Tobagi, “A constant quality MPEG-1 video encoding scheme and its traffic characterization,’’ in Proc. Int. Picture Coding Symp., Mar. 1996, vol. 1, pp. 105–110.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[7] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, “Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization,” *IEEE Trans. Image Process.*, Vol. 22, no. 4, pp.1418-1429, 2013.

[8] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, “Perceptual rate distortion optimization using structural similarity index as quality metric,” *IEEE Trans. Circuits Syst. Video Technol.,* vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[9] C. Yeo, H. Tan, Y. Tan, “SSIM-based adaptive quantization in HEVC”, IEEE International Conference on Speech and Signal Processing (ICASSP), pp.1690-1694, 2013.

[10] S. Wang, “Study on Video Coding Techniques Based on Visual Charactistics,” Doter degree thesis, 2013, Peking University.

[11] D. J. Heeger. “Normalization of cell responses in cat striate cortex,” *Visual Neuroscience*, vol.9, no.2, pp.181-197, 1992.

[12] A. Rehman and Z. Wang, “Reduced-reference image quality assessment by structural similarity estimation,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3378-3389, Aug. 2012.

[13] M. J. Wainwright and E. P. Simoncelli, “Scale mixtures of gaussians and the statistics of natural images,” *Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 855–861, 2000.

[14] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[15] T. Wiegand and B. Girod, “Lagrange multiplier selection in hybrid video coder control,” in Proc. Int. Conf. Image Process. (ICIP), pp. 542–545, 1998.

[16] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullian, “Rate constrained coder control and comparison of video coding standards,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[17] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, “Perceptual rate distortion optimization using structural similarity index as quality metric,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1614–1624, Nov. 2010.

[18] T. Wiegand and B. Girod, “Lagrange multiplier selection in hybrid video coder control,” in Proc. Int. Conf. Image Process., 2001, pp. 542–545.1998.

[19] G. Bjontegaard, “Calculation of average PSNR difference between RD-curves,” in Proc. ITU-T Q.6/SG16 VCEG 13th Meeting, Austin, TX, Apr. 2001, document VCEG-M33.