

# Rate-Distortion Optimization Using Adaptive Lagrange Multipliers

Fan Zhang<sup>IP</sup>, *Member, IEEE*, and David R. Bull, *Fellow, IEEE*

**Abstract**—In current standardized hybrid video encoders, the Lagrange multiplier determination model is a key component in rate-distortion optimization. This originated some 20 years ago based on an entropy-constrained high-rate approximation and experimental results obtained using an H.263 reference encoder on limited test material. In this paper, we present a comprehensive analysis of the results of a Lagrange multiplier selection experiment conducted on various video content using H.264/AVC and HEVC reference encoders. These results show that the original Lagrange multiplier selection methods, employed in both video encoders, are able to achieve optimum rate-distortion performance for I and P frames, but fail to perform well for B frames. The relationship is identified between the optimum Lagrange multipliers for B frames and distortion information obtained from the experimental results, leading to a novel Lagrange multiplier determination approach. The proposed method adaptively predicts the optimum Lagrange multiplier for B frames based on the distortion statistics of recent reconstructed frames. After integration into both H.264/AVC and HEVC reference encoders, this approach was evaluated on 36 test sequences with various resolutions and differing content types. The results show consistent bitrate savings for various hierarchical B frame configurations with minimal additional complexity. BD savings average approximately 3% when constant quantization parameter (QP) values are used for all frames, and 0.5% when non-zero QP offset values are employed for different B frame hierarchical levels.

**Index Terms**—Video compression, rate-distortion optimization, Lagrange multiplier determination, H.264/AVC, HEVC.

## I. INTRODUCTION

VIDEO compression, has been a key enabler for video storage, conferencing, broadcasting and streaming since the early 1980s, when the first widely adopted international coding standard, H.261 [1], was established. With recent advances in video and communication technologies, the demand for video content is ever increasing, with 73% of all Internet bandwidth consumed by video in 2016. This figure is predicted to increase to 82% in 2021 [2].

The latest video compression standard, High Efficiency Video Coding (HEVC) [3], offers improved compression

Manuscript received July 21, 2016; revised December 4, 2016, February 25, 2017, and April 24, 2018; accepted September 19, 2018. Date of publication October 5, 2018; date of current version October 2, 2019. This work was supported by EPSRC under Grant EP/J019291/1. This paper was recommended by Associate Editor M. Flierl. (*Corresponding author: Fan Zhang.*)

The authors are with the Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: fan.zhang@bristol.ac.uk; dave.bull@bristol.ac.uk).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2018.2873837

performance over its predecessors, especially on higher resolution video content. This improvement is due to the introduction of new coding tools, such as more flexible macroblock sizes for prediction and transformation, finer intra prediction modes, improved de-blocking and loop filters, and enhanced interpolation in motion compensation. However the rate-distortion optimization (RDO) module in the HEVC reference encoder (and also that used as the basis for many actual HEVC deployments) employs a model almost identical to those used in most video encoders since H.263 [4].

The RDO approach employed in both HEVC and H.264/AVC [5] reference software (although non-normative) is based on an entropy-constrained high-rate approximation [6], [7]. This method formulates the coding parameter selection problem as finding the minimum of a Lagrange cost function, trading off rate ( $R$ ) and distortion ( $D$ ), and exploits the relationship between  $R$  and  $D$  using an approximate logarithmic function with constant parameters [8]. Model parameters were determined based on I and P frame coding results for three sequences using H.263, and were reported to be content independent [8]. As coding tools have advanced, especially with the frequent use of bi-directional inter prediction and different referencing structure in modern encoders, the optimality of this model has not been fully re-assessed.

In this paper, we address three specific research questions in order to improve the Lagrange multiplier ( $\lambda$ ) determination method in RDO:

- 1) Optimality – does the RDO model employed by modern hybrid encoders still provide optimum rate distortion (RD) performance for all I, P, and B frames?
- 2) Independence – are the optimum Lagrange multipliers still approximately constant across various video content with identical quantization parameters (QP)?
- 3) Predictability – if a negative answer is observed in question 2), can any video features or coding statistics be used to predict the optimum Lagrange multipliers?

This paper provides a comprehensive extension of our previous work in [9], where the Lagrange multiplier selection approach was originally introduced, applied to a simple Group of Pictures (GOP) structure (GOP length 4 with non-hierarchical B frames).<sup>1</sup> We explore the answers to the questions proposed above, and an adaptive  $\lambda$  determination model, extended from that in [9], is presented for application

<sup>1</sup>We define GOP length in this paper as the number of successive bi-directionally predicted B frames plus one.

scenarios with various GOP lengths. Both hierarchical and non-hierarchical GOP structures are employed to test the performance of this approach.

The remainder of this paper is structured as follows. Section II describes the RDO problem and outlines some of the most influential Lagrange multiplier determination models. The experiment on multiplier selection and its results are presented in Section III. In Section IV, a content-based adaptive Lagrange multiplier determination method is proposed, while the evaluation results of this approach are reported and discussed in Section V. Finally, Section VI provides conclusions and implications on future research directions.

## II. BACKGROUND

This section is divided into three subsections. The first introduces the rate distortion optimization (RDO) problem in the context of hybrid video encoders. Previous research work on RDO modeling is then briefly reviewed, followed by a description of the Lagrange multiplier determination models employed in both H.264/AVC and HEVC reference encoders.

### A. The Rate Distortion Optimization Problem

Typically, hybrid video encoders select optimum coding parameters  $\mathbf{p}_{\text{opt}}$  by minimizing a Lagrange cost function of rate  $R$  and distortion  $D$  [10]:

$$\min\{J\}, \quad \text{where } J = D(\mathbf{p}) + \lambda \cdot R(\mathbf{p}), \quad (1)$$

where  $\mathbf{p}$  is the vector of coding parameters including prediction modes, block partitions, etc., and  $\lambda$  represents the Lagrange multiplier. This optimization process is iterated during the compression process at various block levels for different types of frame.

Finding the minimum of a function is a common problem in calculus [11]. In our case this can be solved, when the cost function  $J$  is a convex function of  $\mathbf{p}$ , and both  $R$  and  $D$  are continuous and differentiable everywhere [12]. The Lagrange multiplier  $\lambda$  can then be derived by setting the derivative of  $J$  to zero. Then:

$$\lambda = -\frac{\partial D}{\partial R}. \quad (2)$$

In order to determine  $\lambda$ , the RD curve should be known beforehand. However, this leads to a chicken and egg problem – it is, in general, difficult to predict accurate RD characteristics of videos before encoding them. In the literature, various solutions have been proposed to solve this problem and these are discussed below.

### B. Rate Distortion Modeling

One important group of rate distortion models are based on the distribution of transformed residuals. For example, the generalized Gaussian distribution based RD models, presented in [13]–[15], demonstrate the precision on modeling residual energy. This type of model is however only appropriate for two pass compression, since its model parameters are content dependent and they can only be determined after the first pass. RD models based on the Cauchy distribution [16] have been

proposed which overcome this shortcoming, providing more accurate estimation of transform residuals. It should be noted however that the parameter determination of Cauchy distribution based RDO is difficult due to the diverging characteristic of the model statistics. The Laplace distribution is considered as a specific case of a generalized Gaussian distribution, and RD models based on this [12], offer a trade-off between prediction accuracy and algorithm complexity. Methods in this class also include  $\rho$ -domain algorithms [17].

Another group of RDO methods employ heuristic approaches to estimate Lagrange multipliers [12]. Typical examples of these include methods based on bitrate statistics [18], [19] and local context [20]. However this type of approach sometimes fails to perform well due to inaccurate empirical RD models.

Alongside advances in quality assessment [21]–[23], perceptual video compression algorithms [24]–[26] have been presented that achieve improved rate quality performance [27]. The structural similarity index (SSIM) [21] is one of the most commonly used quality assessment methods for in-loop rate quality optimization (RQO) due to its efficiency and simplicity. Recent work, [28]–[30], has demonstrated the rate quality performance improvement possible with SSIM-based RQO when compared to conventional RDO approaches. It should be emphasized that RQO inspired video compression is still in its infancy, and quality metrics with lower computational complexity (e.g. SSIM) do not always correlate well with subjective quality opinions [23]. In contrast, more advanced methods, such as MOVIE [22] and STMAD [31], are not appropriate for in-loop application due to their high complexity and/or latency characteristics. More recent contributions, such as PVM (Perception-based Video Metric) [23], offer the potential for lower latency and complexity, but are still immature in this respect.

In the context of the above discussion, our focus in this paper is on the enhancement of existing rate-distortion optimization methods using mean squared error (MSE) to assess video quality. While the use of perceptual metrics may provide more robustness in the future, this approach is not applicable at the present time due to the complexity and consistency issues associated with existing metrics.

### C. RDO in H.264/AVC and HEVC Reference Encoders

The RD model most commonly used in modern hybrid video encoders was proposed by Sullivan and Wiegand [8] for entropy-constrained quantization based on a high rate approximation [6], [7], where  $R$  is formulated as the logarithmic function of  $D$ ,

$$R(D) = a \cdot \log_2 \left( \frac{b}{D} \right), \quad (3)$$

where  $a$  and  $b$  are two parameters characterizing the relationship between  $R$  and  $D$ . According to the high rate approximation, the distortion  $D$  can be modeled using the quantization interval  $Q$  as:

$$D = Q^2/3, \quad (4)$$

where  $Q$  can be obtained from the quantization parameter (QP) in H.264 or HEVC using:

$$Q^2 = 2^{(QP-12)/3}. \quad (5)$$

If (3)-(5) are substituted into (2), it provides the result:

$$\lambda = c \cdot 2^{(QP-12)/3}, \quad (6)$$

in which  $c = \ln 2 / (3a)$ .

In order to determine the value of  $c$ , Sullivan and Wiegand [8] conducted a Lagrange multiplier selection experiment on three video sequences using an H.263 reference encoder. The experimental results show that the parameter  $c$  is approximately independent of video content, with a fixed value of 0.85 for inter frames.

Extended from this model, the Lagrange multiplier determination approaches employed in H.264 and HEVC reference encoders (JM and HM respectively), as described in (7) and (8), have been developed with the consideration of bi-directional inter frames. In this paper, we follow the same definitions of I, P and B frames as in the H.264/AVC [5] and HEVC [32] standards. For clarity, we further adopt the following definitions here.  $B_p$  frames are defined as B frames which are inter-predicted only from temporally previous frames. Frames using both previous and subsequent reference frames are defined as  $B_b$  frames.

$$\begin{cases} \lambda_{JM,I} = 0.57 \cdot 2^{(QP-12)/3} \\ \lambda_{JM,P} = 0.85 \cdot 2^{(QP-12)/3} \\ \lambda_{JM,B} = 0.68 \cdot \max(2, \min(4, (QP - 12)/6)) \cdot 2^{(QP-12)/3}, \end{cases} \quad (7)$$

and

$$\begin{cases} \lambda_{HM,I} = (1 - \max(0, \min(0.5, 0.05 N_B))) \cdot 0.57 \cdot 2^{(QP-12)/3} \\ \lambda_{HM,P/B_p} = f \cdot 2^{(QP-12)/3} \\ \lambda_{HM,B_b} = f \cdot \max(2, \min(4, (QP - 12)/6)) \cdot 2^{(QP-12)/3}. \end{cases} \quad (8)$$

In equation (8),  $f$  is referred to as the ‘QPfactor’ in the HEVC HM reference encoder, having a default value of 0.5. The ‘QPfactor’ can be configured differently for frames at various temporal layers in a GOP [33], [34].  $N_B$  is the number of consecutive  $B_b$  frames in a GOP. The Lagrange multiplier model for the HEVC reference encoder has been modified according to the recent recommended configurations in JCTVC-X0038 [35] (using ‘QPfactor’ value of 1 and larger ‘QPoffset’ for each hierarchical B frames) to achieve improved rate distortion performance. It is noted that, in H.264/AVC and HEVC reference encoders, the Lagrange multiplier is modeled as a function of QP, and is independent of video content. Alternative solutions, such as [36], have also been proposed using a fixed Lagrange multiplier to determine QPs for frames at various hierarchical levels.

### III. AN EXPERIMENT ON $\lambda$ SELECTION

It was noted in Section II that the Lagrange multiplier determination methods used in H.264/AVC and HEVC reference encoders employ a basic model whose parameter was empirically derived in [8] using an H.263 reference encoder. The optimality of this model has not been fully validated on modern video encoders using different referencing structures, which can lead to significant changes in RD characteristics.

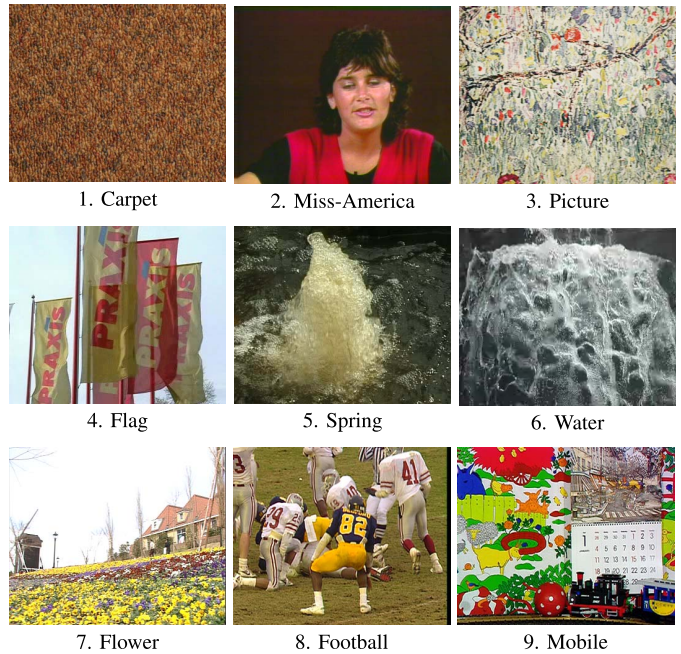


Fig. 1. Sample frames from test sequences used in the  $\lambda$  selection experiment.

TABLE I  
THE VIDEO DATASET USED IN THE  $\lambda$  SELECTION EXPERIMENT

Class	A	B	C
Sequence	1. Carpet 2. Miss-America 3. Picture	4. Flag 5. Spring 6. Water	7. Flower 8. Football 9. Mobile
Source	BVI & Standard	DynTex	Standard & DynTex

#### A. Experimental Methodology

In order to investigate the optimality of the  $\lambda$  determination approaches, we conducted a Lagrange multiplier selection experiment comparing the RD performance using various test multiplier values  $\lambda_{\text{test}}$  with that using the corresponding original multipliers  $\lambda_{\text{orig}}$ , derived from (7) and (8).

The range of  $\lambda_{\text{test}}$  is given by:

$$0.2 \leq \frac{\lambda_{\text{test}}}{\lambda_{\text{orig}}} \leq 5. \quad (9)$$

This experiment was conducted using various test material at CIF ( $352 \times 288$ ) resolution (YUV 4:2:0).<sup>2</sup> In total nine sequences from DynTex [37], the BVI texture database [38], and standard test sequence pools were employed. This dataset was further divided into three classes according to dominant video content: (A) slow movement videos, (B) dynamic texture clips, and (C) mixed content. TABLE I provides a list of these sequences, while Fig. 1 shows their sample frames.

In this experiment, five subtests were conducted with different objectives. The first two assessed the optimality of the

<sup>2</sup>Based on computational complexity consideration, low resolution (CIF) videos were used during the training stage. Similar parameters were obtained when a limited number of higher resolution sequences ( $1920 \times 1080$ ) were used. The performance of the proposed algorithm on a wide range of formats (including nine HD sequences) is fully presented in Section V.

TABLE II  
THE GOP SETTINGS OF THE CONDUCTED  $\lambda$  SELECTION TESTS

Test No.	i	ii	iii	iv	v
Objectives	I frame	P/B <sub>p</sub> frame	B <sub>b</sub> frame		
No. of frames tested	100	100	101	121	129
No. of I frames	100	1	1	1	1
No. of P/B <sub>p</sub> frames	0	99	25	15	8
No. of B <sub>b</sub> frames	0	0	75	105	120
GOP length	1	1	4	8	16

model for I and P/B<sub>p</sub> frames respectively, while the last three investigated that for B<sub>b</sub> frames with different GOP structures. Note that we only modified the Lagrange multipliers for the tested frame types. The GOP configurations for these five tests are given in TABLE II.

JM 15.1 and HM 14.0 were used for H.264 and HEVC respectively; identical QP values were employed for all types of frames; the range of tested QP values was from 27 to 42 with an interval of 5; Main profile and non-hierarchical B frames were selected for JM; Main profile and hierarchical B frames were tested for HM.

It should be noted that constant QP values (QP offset equals zero) are used for all frames in this experiment for the HEVC HM encoder. This differs from the recommended configurations in [33], [35], and [39], where fixed QP offset values are used for different hierarchical B frame levels to improve overall rate-distortion (R-D) performance. Based on the recent work in [40], using constant QP offset values does not always offer optimum R-D performance for all types of content, and QP offset values in the HEVC reference encoder should be adapted based upon video content. Since the purpose of this paper is solely to investigate the influence of Lagrange multipliers on R-D performance, constant QP values are employed in our training process, as this eliminates the confounding influence of QP offset.

### B. Results for I and P/B<sub>p</sub> Frames

The performance of the Lagrange multiplier determination methods for I and P/B<sub>p</sub> frames is shown in Fig. 2.(a-d), where original multiplier values ( $\lambda_{orig}$ ) are plotted alongside the optimum ones ( $\lambda_{opt}$ ). These optimum Lagrange multipliers were selected to have the best overall RD performance for all frames compared to the original RD curves generated using  $\lambda_{orig}$ .

It can be observed that  $\lambda_{orig}$  curves associated with I and P/B<sub>p</sub> frames do correlate well with corresponding  $\lambda_{opt}$  values for both H.264 and HEVC encoders, although several outliers exist for the case of P/B<sub>p</sub> frames. Among all 9 test sequences and 4 QP values, only 3  $\lambda_{orig}$  values out of 36 are not able to offer optimum RD performance for H.264 P frame coding, while 4 outliers appear for HEVC P/B<sub>p</sub> frames. This indicates that the original  $\lambda$  determination models used in both encoders perform well for I and P frame encoding.

### C. Results for B<sub>b</sub> Frames

Fig. 2.(e,f) illustrates the test results for B<sub>b</sub> frames with various GOP sizes. The test multiplier values were only applied on

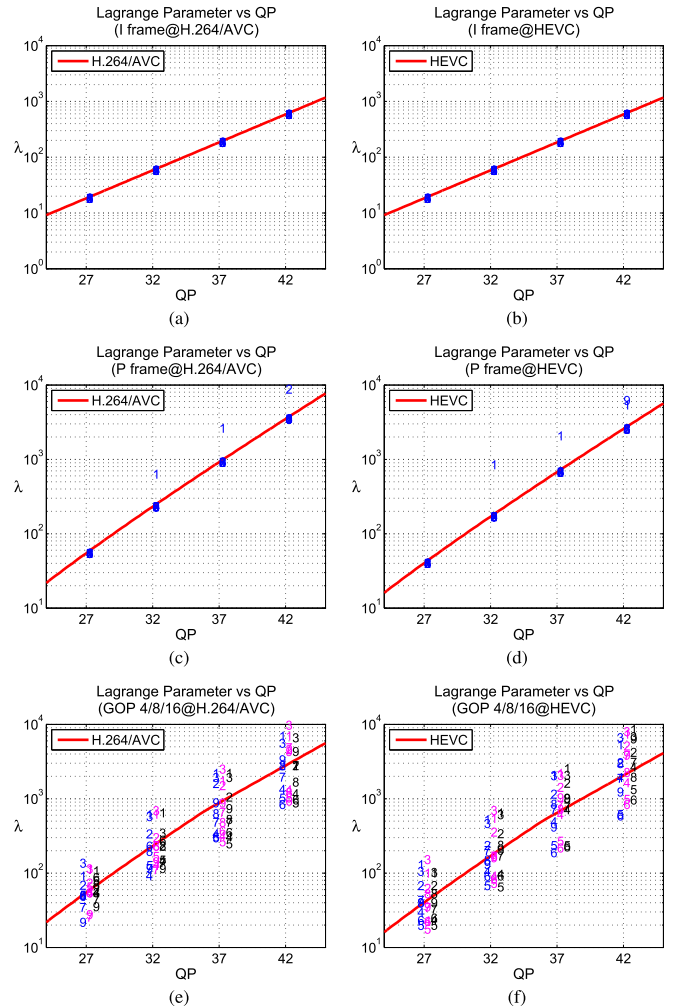


Fig. 2. The optimum Lagrange multipliers ( $\lambda_{opt}$ ) versus corresponding original values ( $\lambda_{orig}$ ). Results for (a) H.264 JM I frame, (b) HEVC HM I frame, (c) H.264 JM P frame, (d) HEVC HM P frame, (e) H.264 JM B frame for GOP length 4, 8 and 16 and (f) HEVC HM B frame for GOP length 4, 8 and 16. The position of each number represents the  $\lambda_{opt}$  values for that sequence at a given QP, while the red curves represent the  $\lambda_{orig}$  values as a function of QP. In subfigures (e) and (f), numbers in blue, pink, and black colors refer to the results for GOP length 4, 8 and 16 respectively.

B<sub>b</sub> frames, which use both temporally previous and subsequent frames as references for inter-prediction. In these cases,  $\lambda_{orig}$  values fail to correlate well with  $\lambda_{opt}$  for both HEVC and H.264 regardless of whether the GOP length is 4, 8 or 16. The failure becomes more evident for static scene content (Class A) and dynamic textures (Class B). These results confirm our conjecture in Section I, that the conventional RDO module does need to be improved for modern video encoders.

As a result of this model failure, the distortion difference between B<sub>b</sub> and P/B<sub>p</sub> frames varies among test sequences. Here we define the ratio between the mean squared error (MSE) of P/B<sub>p</sub> frames ( $MSE_P$ ) and that of B<sub>b</sub> frames ( $MSE_B$ ) as follows (only Y components of reconstructed and original frames are used for calculating MSE).

$$r_{MSE} = MSE_P / MSE_B. \quad (10)$$

In order to investigate the relationship between  $r_{MSE}$  and the mismatch between  $\lambda_{opt}$  and  $\lambda_{orig}$  for B<sub>b</sub> frames, a second

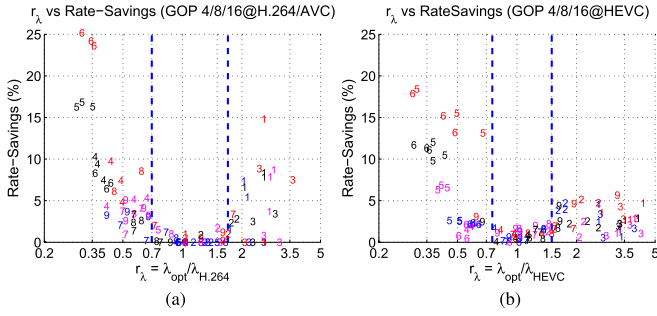


Fig. 3. The bitrate savings at various  $r_\lambda$  for GOP length 4, 8 and 16. These are based on the results for all frames. Results for (a) H.264 and (b) HEVC. The position of each number represents the bitrate saving at the corresponding  $r_\lambda$  ratio for that sequence.

ratio is defined as:

$$r_\lambda = \lambda_{opt}/\lambda_{orig}. \quad (11)$$

Fig. 4 demonstrates the relationship between  $r_{MSE}$  and  $r_\lambda$  for  $B_b$  frame coding with various GOP structures. It can be seen that the scatter plots for GOP length 4 using H.264 and HEVC encoders both fit well to a power function, and those for GOP 8 and 16 follow similar fitting curves as GOP 4 cases, only with a shift to the left. Based on this observation, we employ a four parameter power function to fit the correlation between  $r_\lambda$  and  $r_{MSE}$  for all cases, as given below:

$$r_{\lambda,fit} = f(r_{MSE}) = a(r_{MSE} + d)^b + c. \quad (12)$$

Here  $a$ ,  $b$ ,  $c$  and  $d$  are parameters which are determined using the dataset in TABLE I. Values of parameter  $a$ ,  $b$  and  $c$  are 2.696, 10.06, and 0.367 for H.264 JM, and 2.197, 5.196, and 0.308 for HEVC HM. Parameter  $d$  is related to GOP sizes. For the three tested scenarios (GOP length 4, 8 and 16), the corresponding  $d$  values are 0, 0.05 and 0.07 for H.264 JM, and 0, 0.04, and 0.1 for HEVC HM.

The overall bitrate saving at each  $\lambda_{opt}$  for  $B_b$  frames over the original RD curve using the corresponding  $\lambda_{orig}$  value for three tested GOP settings is illustrated in Fig. 3. The savings are content dependent and vary from 0% to 25% for H.264 and from 0% to 18% for HEVC. It can also be clearly seen that, for both encoders and all three GOP lengths, bitrate savings are below 2% if  $r_\lambda$  falls within the range between the two blue dotted lines.

#### D. Summary

In summary, our Lagrange multiplier selection experiment assessed the optimality of existing  $\lambda$  models in both H.264 and HEVC reference encoders, and we explored the answers to the questions proposed in Section I. Based on the experimental results above, four important findings are summarized as below.

- 1) Existing Lagrange multiplier determination models in both H.264 and HEVC reference encoders are close to optimum for I and P/B<sub>p</sub> frames, but do not perform well for B<sub>b</sub> frames.

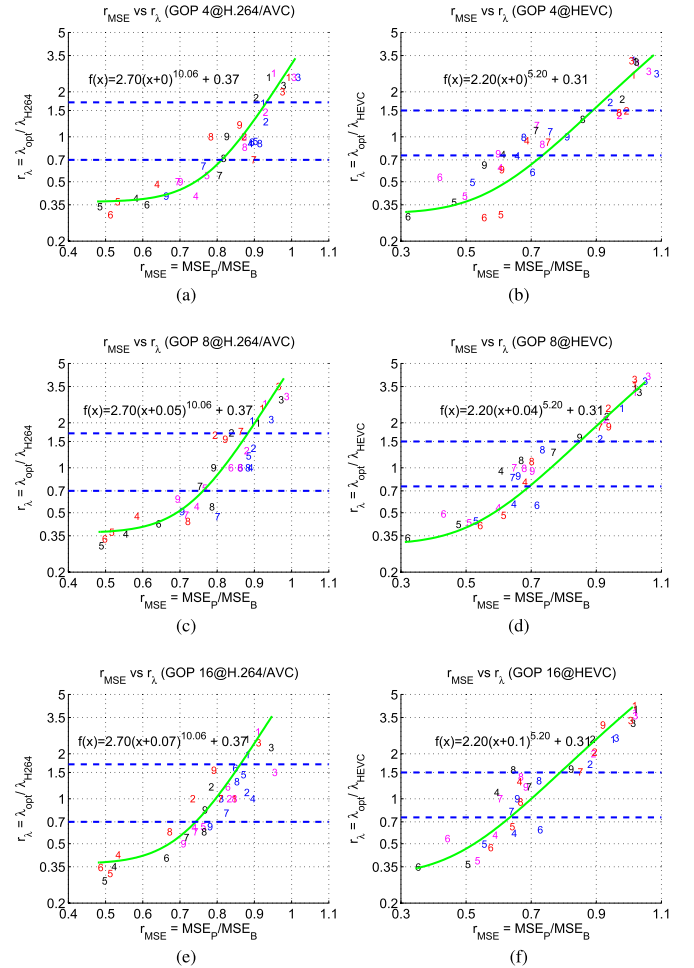


Fig. 4. The  $r_{MSE}$  versus  $r_\lambda$ . Green curves represent the fitting curves. Results for (a) H.264 with GOP length 4, (b) HEVC with GOP length 4, (c) H.264 with GOP 8, (d) HEVC with GOP 8, (e) H.264 with GOP 16 and (f) HEVC with GOP 16.

- 2) Optimum  $\lambda$  values for  $B_b$  frames in both encoders are content dependent – higher for static scenes and lower in cases with significant dynamic content.
- 3) Distortion statistics could be used to predict optimum Lagrange multipliers for  $B_b$  frames.

#### IV. PROPOSED ALGORITHM

In order to adaptively predict optimum Lagrange multipliers for  $B_b$  frames, a novel content-based determination approach is proposed, inspired by the experimental results in Section III and our preliminary model in [9], which uses lower Lagrange multiplier values for dynamic scenes, and higher ones for static content. This method operates under the assumption that *within a few temporally localized frames, providing there are no significant content changes, the RD characteristics are approximately uniform*. Lagrange multipliers could thus be adaptively modified according to distortion statistics from recently encoded frames. This assumption may of course break when there are scene cuts. To account for this, a shot cut detector should be employed prior to  $\lambda$  adaptation.

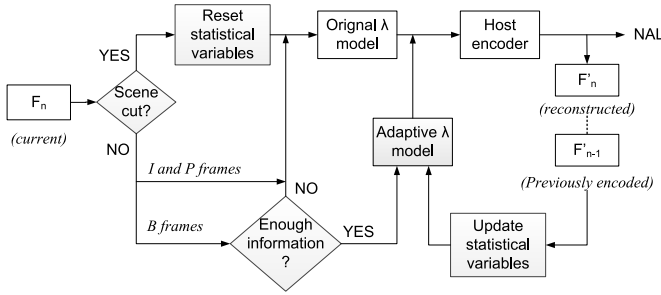


Fig. 5. The proposed method.

A diagrammatic illustration of the proposed method is shown in Fig. 5. Before encoding each frame, possible shot transitions are firstly identified using a scene cut detection approach based on histogram differences. In cases with scene cuts, when the uniform assumption is not applicable, all statistical variables are reset, and the original Lagrange multiplier model is used. For frames without shot transitions, only  $B_b$  frame  $\lambda$  values are adaptively modified if sufficient distortion information has been recorded. The distortion of every encoded frame is recorded to update the statistical variables for future  $\lambda$  modification.

The proposed algorithm consists of three primary sub stages: scene cut detection, distortion information updating and Lagrange multiplier modification. These are described in detail below.

#### A. Scene Cut Detection

Scene cut detection can be based on numerous measures including histogram differences (HistD), edge change ratio, and sum of absolute differences [41]. In the context of video compression, we employ a simple but efficient HistD-based approach with a constant threshold.

In our method, the normalized luma histogram of the current frame,  $\text{Hist}_t$ , is firstly computed alongside that of its previous coded frame (if applicable),  $\text{Hist}_{t_p}$ . Their average absolute difference  $\text{HistD}_t$  is then compared with a fixed threshold  $\text{TH}_{\text{SC}}$  to identify the scene cut. This process is described by (13) and (14)

$$\text{HistD}_t = \sum_{i=0}^{2^L-1} \frac{|\text{Hist}_t(i) - \text{Hist}_{t_p}(i)|}{2^L}, \quad (13)$$

and

$$\begin{cases} \text{HistD}_t \geq \text{TH}_{\text{SC}}, \text{ there is scene cut in this frame} \\ \text{HistD}_t < \text{TH}_{\text{SC}}, \text{ there is no scene cut in this frame.} \end{cases} \quad (14)$$

where  $L$  represents bit depth, and  $\text{TH}_{\text{SC}}$  is chosen as 0.002 for normalized histogram of 8 bit luminance. This value is empirically obtained through a preliminary training process on limited sequences, and it was found not to be significantly sensitive to content type.

As shown in Fig. 5, when a scene cut is detected, all existing statistical variables are reset, and the original Lagrange multipliers for this frame will be used in rate-distortion optimization.

#### B. Distortion Information Updating

To adaptively adjust Lagrange multipliers, sufficient distortion information – from at least one consecutive GOP must be recorded. Here two distortion statistics,  $D_P$  and  $D_B$ , are defined for  $P/B_p$  and  $B_b$  frames respectively.

$$\begin{cases} D_{P/B,k} = \text{MSE}_{P/B,k}, \text{ if } D_{P/B,k-1} = 0 \text{ or } k = 1 \\ D_{P/B,k} = \theta_1 \cdot D_{P/B,k-1} + \theta_2 \cdot \text{MSE}_{P/B,k}, \text{ otherwise,} \end{cases} \quad (15)$$

in which  $D_{P/B,k}$  represents the accumulated distortion  $D_P$  or  $D_B$  based on frame type, having its initial value set to zero.  $k$  is the number of  $P/B_p$  or  $B_b$  frames which have been coded, which is counted following the encoding order.  $\text{MSE}_{P/B,k}$  stands for the mean squared error of the most recently coded frame.  $\theta_1$  and  $\theta_2$  are pre-configured weighting parameters, combining the existing distortion with the latest MSE. Here  $\theta_1 + \theta_2 = 1$ , and  $\theta_2 > \theta_1$ . This configuration is to place greater emphasis on recently encoded frames.

#### C. Lagrange Multiplier Modification

With sufficient distortion information recorded from previously coded frames, the Lagrange multiplier for the current  $B_b$  frame ( $\lambda_n$ ) is adaptively modified from that of the most recently coded  $B_b$  frame ( $\lambda_{n-1}$ ) following:

$$\lambda_n = \lambda_{n-1}(a(r_{\text{MSE}} + d)^b + c), \quad (16)$$

where  $r_{\text{MSE}}$  is the distortion ratio which is derived as follows:

$$r_{\text{MSE},n} = \frac{D_{P,m}}{D_{B,n}}, \quad (17)$$

in which  $m$  represents the number of encoded  $P/B_p$  frames when  $n$   $B_b$  frames have been processed.

It is noted that the proposed model predicts the optimum Lagrange multiplier values based on the distortion statistics of previously encoded frames rather than those for the current frame. This may lead to a slightly inaccurate estimation, when the rate-distortion performance between frames is not identical. This inaccuracy can be avoided if the Lagrange multiplier is only modified when  $a(r_{\text{MSE}} + d)^b + c$  becomes significantly different from 1. In Section III-C, the bitrate savings obtained using optimum Lagrange multipliers were observed to become less significant (below 2%) when  $r_\lambda$  falls within a certain range ( $r_1, r_2$ ). We thus exploit this observation in our approach, keeping the modified Lagrange multiplier constant:

$$\lambda_n = \lambda_{n-1}, \text{ if } r_1 < \frac{\lambda_n}{\lambda_{n-1}} < r_2. \quad (18)$$

In cases when there is significant difference between  $\lambda_n$  and  $\lambda_{n-1}$ , we confine this change to within a  $\pm 5\%$  range to avoid noticeable quality variations, i.e.

$$\begin{cases} \lambda_n = 95\% \cdot \lambda_{n-1}, \text{ if } \lambda_n/\lambda_{n-1} < 95\% \\ \lambda_n = 105\% \cdot \lambda_{n-1}, \text{ if } \lambda_n/\lambda_{n-1} > 105\%. \end{cases} \quad (19)$$

Using this adaptive algorithm, Lagrange multipliers for all  $B_b$  frames can be iteratively obtained, and used for mode, partition and prediction selections at various levels in the rate-distortion optimization process.

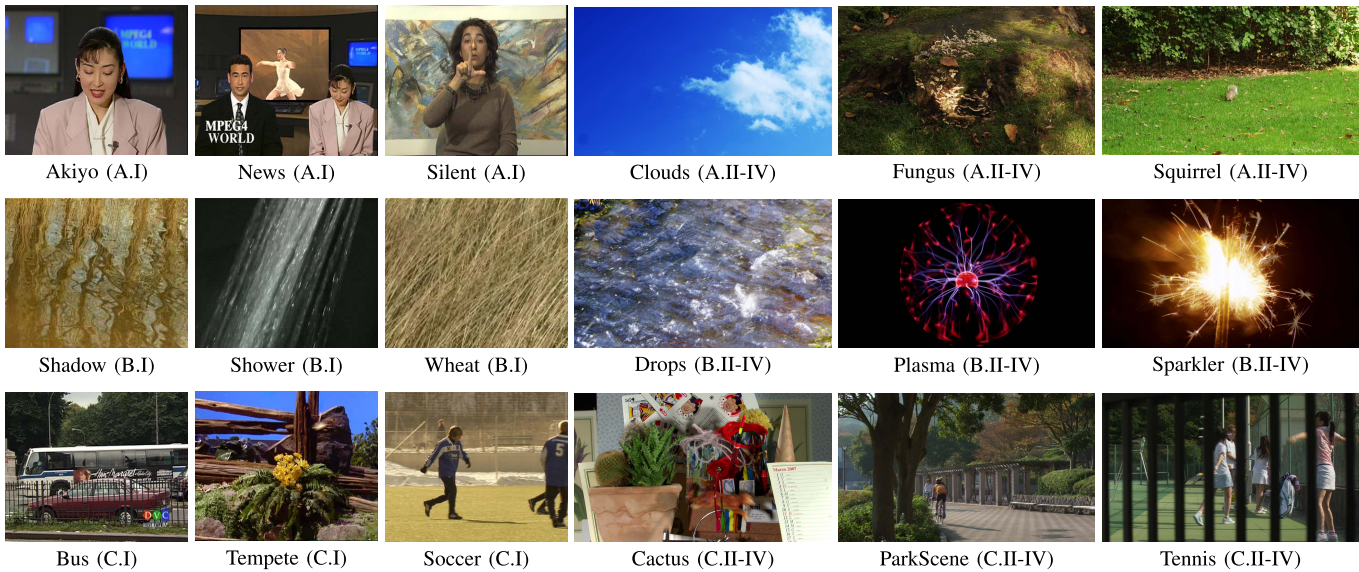


Fig. 6. Sample frames from test sequences along with their class/group indices.

TABLE III

MODEL PARAMETERS USED IN THE PROPOSED  $\lambda$  DETERMINATION METHOD FOR BOTH H.264/AVC AND HEVC REFERENCE ENCODERS

Encoder	$a$	$b$	$c$	$d$		
				GOP 4	GOP 8	GOP 16
H.264 JM	2.696	10.06	0.367	0	0.05	0.07
HEVC HM	2.197	5.196	0.308	0	0.04	0.01

Encoder	$r_1$	$r_2$	$\theta_1$	$\theta_2$	$TH_{SC}$
H.264 JM	0.7	1.7	0.2	0.8	0.002
HEVC HM	0.75	1.5	0.2	0.8	0.002

#### D. Model Parameters

There are in total nine parameters employed in our adaptive algorithm:  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $r_1$ ,  $r_2$ ,  $\theta_1$ ,  $\theta_2$  and  $TH_{SC}$ . The former three are obtained based on the power function fitting described in Section III-C. Parameter  $d$  is related to the used GOP sizes, and its values for GOP length 4, 8 and 16 are also determined based on the fitting. These three GOP lengths are commonly used in both H.264 and HEVC reference encoders. For other GOP configurations, the value parameter  $d$  may vary and could be obtained using the same approach.

Moreover, thresholding parameters  $r_1$  and  $r_2$  are configured based on the experimental results in Fig. 3. Finally  $\theta_1$  and  $\theta_2$  are weighting parameters for updating distortion information. All parameter values are listed in TABLE III.

### V. RESULTS AND DISCUSSION

After integration into the H.264/AVC and HEVC reference software, the proposed adaptive Lagrange multiplier determination method was tested on a video dataset with various content at different resolutions. The RD performance of the proposed method is compared with that of the original  $\lambda$  determination model in both encoders under multiple test conditions. The computational complexity of this approach is also estimated.

TABLE IV

TEST CLIPS USED FOR EVALUATION COMPRESSION PERFORMANCE

Group & Class	Sequence and Length	Source
I.A II-IV.A	Akiyo (300f), News (300f), Silent (300f) Clouds (240f), Fungus (240f), Squirrel (240f)	Standard BVI
I.B II-IV.B	Shadow (300f), Shower (250f), Wheat (250f) Drops (300f), Plasma (240f), Sparkler (300f)	DynTex BVI
I.C II-IV.C	Bus (150f), Tempete (260f), Soccer (300f) Cactus (300f), ParkScene (240f), Tennis (240f)	Standard HEVC

#### A. Test Dataset

Thirty-six test clips are used, all in progressive YUV 4:2:0 format, obtained from public video databases including the HEVC recommended test pool [39], [42], the DynTex database [37], the BVI video texture dataset [38] alongside other commonly used sequences. These test sequences can be divided into three content classes: (A) static scenes, (B) dynamic scenes, and (C) mixed scenes, as in Section III. Videos in each class can be further classified into four groups according to their spatial resolutions: three at CIF ( $352 \times 288$ ) resolution, three at  $416 \times 240$ , three at  $832 \times 480$  and three at  $1920 \times 1080$ . The latter three groups contain videos at different spatial resolutions with identical content, in order to investigate the influence of various resolutions. A description of these videos is provided in Table IV, and their sample frames are shown in Fig. 6.

In order to quantify the content of this dataset, three low-level feature descriptors were computed for each original video: mean spatial information (SI), colorfulness (CF) and mean temporal information (TI). The detailed description of these features can be found in [43] and [44]. The coverage and distribution of these features on the test dataset are shown in Fig. 7. It is noted that this dataset offers good coverage over

TABLE V  
SUMMARY OF COMPRESSION RESULTS

Observed Frames		All frames				B <sub>0</sub> frames only			
Anchor Encoders		H.264 JM		HEVC HM		H.264 JM		HEVC HM	
Test No.	Group/Class	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
Test i. (4-HB)	I	0.21dB	-3.4%	0.13dB	-2.5%	0.28dB	-12.3%	1.31dB	-15.3%
	II	0.18dB	-2.9%	0.13dB	-2.8%	0.67dB	-11.0%	1.07dB	-18.9%
	III	0.23dB	-4.2%	0.21dB	-4.3%	0.26dB	-4.9%	0.72dB	-13.5%
	IV	0.17dB	-3.1%	0.10dB	-2.4%	0.51dB	-9.1%	0.99dB	-19.2%
	A	0.11dB	-2.3%	0.19dB	-4.2%	1.19dB	-26.7%	2.73dB	-36.8%
	B	0.45dB	-7.1%	0.17dB	-3.1%	0.10dB	-1.2%	0.24dB	5.1%
	C	0.03dB	-0.7%	0.06dB	-1.8%	-0.01dB	-0.0%	0.59dB	-18.4%
Overall		0.20dB	-3.4%	0.14dB	-3.0%	0.43dB	-9.3%	1.02dB	-16.7%
Test ii. (8-HB)	I	0.18dB	-2.9%	0.13dB	-2.5%	0.75dB	-12.2%	1.16dB	-13.1%
	II	0.15dB	-2.5%	0.12dB	-2.6%	0.71dB	-12.1%	0.90dB	-15.2%
	III	0.19dB	-3.4%	0.21dB	-4.3%	0.29dB	-6.4%	0.51dB	-9.7%
	IV	0.15dB	-2.6%	0.10dB	-2.2%	0.55dB	-10.3%	0.93dB	-17.5%
	A	0.14dB	-3.0%	0.23dB	-4.7%	1.43dB	-25.3%	2.32dB	-32.7%
	B	0.36dB	-5.6%	0.18dB	-3.3%	0.26dB	-4.2%	0.05dB	1.2%
	C	-0.00dB	0.0%	0.02dB	-0.7%	0.03dB	-1.2%	0.36dB	-10.1%
Overall		0.17dB	-2.9%	0.14dB	-2.9%	0.57dB	-10.2	0.88dB	-13.9%
Test iii. (16-HB)	I	0.11dB	-1.7%	0.12dB	-2.4%	0.50dB	-8.1%	1.09dB	-12.4%
	II	0.13dB	-2.2%	0.12dB	-2.7%	0.46dB	-8.7%	0.83dB	-14.7%
	III	0.12dB	-2.0%	0.22dB	-4.6%	0.16dB	-3.3%	0.45dB	-9.6%
	IV	0.12dB	-2.0%	0.09dB	-2.0%	0.41dB	-8.0%	0.67dB	-14.1%
	A	0.16dB	-3.3%	0.27dB	-5.7%	0.96dB	-18.3%	2.06dB	-30.3%
	B	0.23dB	-3.3%	0.11dB	-2.1%	0.20dB	-2.8%	0.02dB	-0.4%
	C	-0.04dB	0.8%	0.04dB	-1.1%	-0.01dB	0.0%	0.20dB	-7.4%
Overall		0.12dB	-1.9%	0.14dB	-3.0%	0.38dB	-7.0%	0.76dB	-12.7%
Test iv. (4-NHB)	I	0.22dB	-3.5%	0.14dB	-2.6%	0.68dB	-12.1%	0.97dB	-12.2%
	II	0.18dB	-2.9%	0.13dB	-2.5%	0.69dB	-12.1%	0.95dB	-14.9%
	III	0.24dB	-4.4%	0.20dB	-4.0%	0.28dB	-6.0%	0.57dB	-9.8%
	IV	0.17dB	-3.1%	0.10dB	-2.2%	0.58dB	-8.6%	0.66dB	-15.2%
	A	0.15dB	-3.1%	0.16dB	-3.3%	1.50dB	-26.2%	2.36dB	-34.5%
	B	0.43dB	-6.7%	0.25dB	-4.5%	0.17dB	-2.7%	0.31dB	6.3%
	C	0.02dB	-0.6%	0.02dB	-0.8%	-0.00dB	-0.2%	0.30dB	-10.8%
Overall		0.20dB	-3.5%	0.14dB	-2.9%	0.56dB	-9.7%	0.79dB	-13.0%
Test v. (8-NHB)	I	0.23dB	-3.8%	0.15dB	-2.9%	0.79dB	-11.8%	0.96dB	-12.9%
	II	0.19dB	-3.1%	0.14dB	-3.1%	0.63dB	-11.0%	0.84dB	-15.8%
	III	0.24dB	-4.3%	0.21dB	-4.4%	0.28dB	-6.0%	0.48dB	-9.2%
	IV	0.19dB	-3.4%	0.12dB	-2.6%	0.53dB	-9.5%	0.69dB	-15.9%
	A	0.20dB	-4.1%	0.27dB	-5.7%	1.33dB	-22.9%	2.10dB	-34.4%
	B	0.42dB	-6.3%	0.18dB	-3.3%	0.33dB	-5.2%	0.05dB	1.0%
	C	0.02dB	-0.5%	0.02dB	-0.8%	0.02dB	-0.7%	0.17dB	-7.0%
Overall		0.21dB	-3.6%	0.16dB	-3.3%	0.56dB	-9.6%	0.74dB	-13.4%
Test vi. (16-NHB)	I	0.22dB	-3.4%	0.16dB	-3.2%	0.57dB	-8.9%	0.83dB	-12.5%
	II	0.19dB	-3.3%	0.16dB	-3.5%	0.48dB	-8.9%	0.78dB	-15.6%
	III	0.21dB	-3.7%	0.21dB	-4.5%	0.21dB	-4.1%	0.41dB	-8.7%
	IV	0.20dB	-3.8%	0.12dB	-2.6%	0.44dB	-8.2%	0.63dB	-15.0%
	A	0.21dB	-4.2%	0.34dB	-7.3%	0.92dB	-16.9%	1.82dB	-32.8%
	B	0.39dB	-5.9%	0.11dB	-2.1%	0.35dB	-5.5%	0.02dB	-0.5%
	C	0.02dB	-0.4%	0.03dB	-1.0%	0.01dB	-0.2%	0.15dB	-5.6%
Overall		0.21dB	-3.5%	0.16dB	-3.5%	0.42dB	-7.5%	0.66dB	-13.0%

N.B. X-HB represents GOP length X with hierarchical B frames, in which X stands for either 4, 8, or 16. X-NHB stands for GOP length X non-hierarchical B frames.

these descriptors, compared with other public video databases reported in [43].

### B. Test Conditions

The proposed algorithm was fully tested under six groups of test condition with different GOP structures (GOP length 4,

8 and 16 with hierarchical and non-hierarchical B frames), as summarized in Table VI. Other primary configuration include: JM 15.1 and HM 14.0 were employed as reference modules for H.264 and HEVC respectively; uniform QPs were used for all test frames – from 22 to 42 with an interval as 5; High profile and Main profile were selected for H.264 and



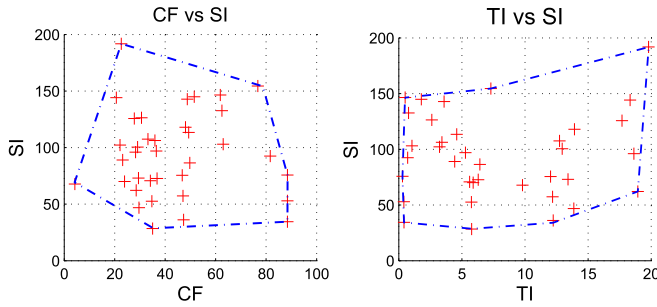


Fig. 7. The feature coverage and distribution of the test dataset.

TABLE VI  
VARIOUS TEST CONDITIONS

Test Conditions	i	ii	iii	iv	v	vi
B frame type	Hierarchical B			Non-hierarchical B		
GOP length	4	8	16	4	8	16

TABLE VII

COMPRESSION RESULTS IN TERMS OF BD-RATE BASED ON PVM

Test	JM	HM	Test	JM	HM
i. (4-HB)	-5.2%	-2.9%	ii. (4-NHB)	-5.4%	-3.1%
iii. (8-HB)	-4.6%	-2.8%	iv. (8-NHB)	-6.2%	-3.2%
v. (16-HB)	-3.6%	-2.5%	vi. (16-NHB)	-6.0%	-3.0%

HEVC encoders respectively; only one I frame was encoded for each test.

The compression performance of the proposed algorithm for both H.264 and HEVC was benchmarked against the corresponding anchor encoders based on the Bjøntegaard delta measurements (BD-rate and BD-PSNR) [45] for the cases (i) all frames (ii) only  $B_b$  frames.

### C. Test Results for Various GOP Structures

The average bitrate savings together with the mean PSNR gains over the anchor encoders are shown in Table V, where the results under various test conditions are provided for all frames and  $B_b$  frames only. For seven test configurations, the average BD-Rate and BD-PSNR values are summarized for four resolution groups (I, II, III, and IV) and three content classes (A, B and C). It can be observed that the proposed method consistently offers superior overall performance for different test groups (resolution) and classes (content). The average bitrate savings for hierarchical B frame structure configurations are approximately 3% over both H.264 and HEVC anchor encoders. It is also noted that this improvement becomes more significant on static and dynamic content (Class A and B) than on video clips with mixed content (Class C) for the various test conditions.

In order to provide some indication of perceptual quality improvements using the proposed method, Table VII shows additional comparative results using the PVM metric [23]. PVM was chosen as it provides improved correlation with

TABLE VIII

HEVC COMPRESSION RESULTS (BD-RATE AGAINST ORIGINAL HEVC HM) BASED ON THE RA CONFIGURATIONS (GOP 8) IN [33] AND [35]

Test Group	I	II	III	IV
JCTVC-L1100	-1.2%	-1.1%	-1.2%	-1.0%
JCTVC-X0038	-0.5%	-0.6%	-0.4%	-0.4%

subjective scores across a wide variety of content types and distortions. The results in Table VII show close agreement with the PSNR-based results in Table V, further validating the benefits of our approach.

Example RD curves comparing the proposed method with conventional H.264 and HEVC reference encoders for three test sequences are shown in Fig. 8. The selected sequences represent typical content from each class based on various test conditions. Evident bitrate savings can be observed from the proposed method over the anchor approach, especially for the ‘Fungus’ sequence in Class A and ‘Shadow’ in Class B.

### D. Test Results on Adaptive QP Configurations for HM

The proposed model was also tested on the HEVC reference encoder using the Random Access (RA) configurations (GOP length 8) in JCTVC-L1100 [33] and JCTVC-X0038 [35], in which fixed non-zero QP offset values are utilized for different B frame hierarchical levels. These configurations have been shown to yield improved overall rate-distortion performance. However they may also produce significant temporal quality variations due to the large QP differences between frames.

The compression results for different test groups (classified by resolution) are summarized in TABLE VIII. It should be noted that our method was trained using a constant QP configuration (QP offset equals 0). Therefore this will clearly not be optimum when large QP offsets are employed. Nevertheless, the proposed approach still shows consistent overall bitrate savings across video groups at various spatial resolutions, with an average BD-rate values at 1.1% and 0.5% for JCTVC-L1100 and JCTVC-X0038 respectively. According to the results in [40], the use of fixed QP offset values does not provide optimum R-D performance for all types of content. More significant bitrate savings may therefore be possible if our Lagrange multiplier determination approach is combined with a content-based adaptive QP model. This is a topic for future research.

### E. Complexity Estimation

Finally, the computational complexity of the proposed algorithm was estimated based on the relative execution time of the proposed and original anchor encoders. The average encoding times for both the proposed method and the anchor were calculated. The results are shown in Table IX which presents the percentage increase in encoding time for the proposed method, referenced against the anchor. The average additional complexity of our approach was found to be insignificant, with only 5% and 2% increases over H.264 JM and HEVC HM modules respectively. The results also indicate that the

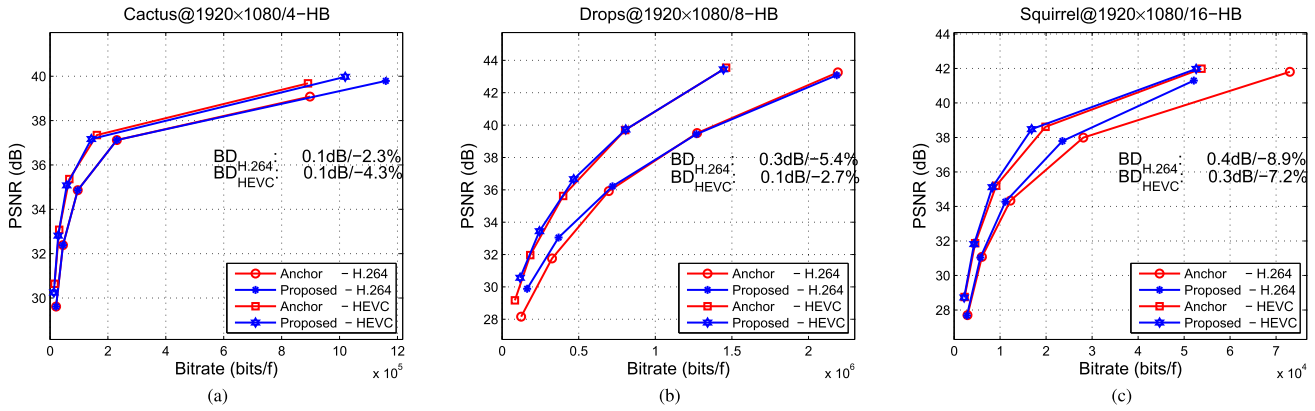


Fig. 8. RD curves for HD sequences Cactus, Drops and Squirrel under various test conditions.

TABLE IX  
COMPLEXITY ANALYSIS

Test	JM	HM	Test	JM	HM
i. (4-HB)	6.6%	2.0%	ii. (4-NHB)	4.6%	1.2%
iii. (8-HB)	6.0%	2.2%	iv. (8-NHB)	2.9%	1.6%
v. (16-HB)	4.2%	1.6%	vi. (16-NHB)	3.3%	1.8%

increases are mainly (more than 90% of the additional complexity) due to the scene cut detection method used, which only consumes linear time  $O(n)$  – where  $n$  is the number of pixels in each frame. All complexity figures were obtained using an Intel Core i7-2600 CPU @3.40Ghz PC platform.

## VI. CONCLUSIONS

In this paper, we conducted a Lagrange multiplier selection experiment using modern hybrid video encoders. Experimental results demonstrate the optimality of existing  $\lambda$  determination methods in H.264/AVC and HEVC reference encoders for encoding I and P/B<sub>p</sub> frames, but highlight the shortcomings of these models for B<sub>b</sub> frames. The relationship between two ratio indices – the distortion ratio between P/B<sub>p</sub> and B<sub>b</sub> frames, and the ratio between the optimum Lagrange multipliers and the original ones was discovered which led to a new adaptive determination method for B<sub>b</sub> frame encoding. This approach has been evaluated for various content types and test conditions. The results show consistent RD performance improvement over the anchor encoders, for both H.264 and HEVC with various hierarchical B-frame configurations. BD-rate savings average 3% when constant QP values are used for all frames, and 0.5% when non-zero QP offset values are employed for different levels in the B-frame hierarchy. In terms of future work, the authors suggest combining adaptive  $\lambda$  determination with varied quantisation parameters, and also performance evaluation using subjective quality assessment.

## REFERENCES

- [1] International Telecommunications Union-Radiocommunication (ITU-R), “Video codec for audiovisual services at p×64 kbit/s,” ITU Recommendation H.261, 1993.
- [2] Cisco, “Cisco visual networking index: Forecast and methodology, 2016–2021,” Cisco, San Jose, CA, USA, White Paper, May 2017. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf>
- [3] 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, Dec. 2012.
- [4] K. Rijkse, “H.263: Video coding for low-bit-rate communication,” *IEEE Commun. Mag.*, vol. 34, no. 12, pp. 42–45, Dec. 1996.
- [5] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), ITU Standard ISO/IEC 14496-10, “Advanced video coding for generic audiovisual services,” ITU-T Rec. H.264, 2005.
- [6] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Upper Saddle River, NJ, USA: Prentice-Hall, 1984.
- [7] H. Gish and N. J. Pierce, “Asymptotically efficient quantizing,” *IEEE Trans. Inf. Theory*, vol. IT-14, no. 5, pp. 676–683, Sep. 1968.
- [8] 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.
- [9] F. Zhang and D. R. Bull, “An adaptive Lagrange multiplier determination method for rate-distortion optimisation in hybrid video codecs,” in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 671–675.
- [10] D. R. Bull, *Communicating Pictures: A Course in Image and Video Coding*. New York, NY, USA: Academic, 2014.
- [11] J. Stewart, *Calculus: Concepts and Contexts, Single Variable*. Pacific Grove, CA, USA: Brooks/Cole, 1998.
- [12] X. Li, N. Oertel, A. Hutter, and A. Kaup, “Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.
- [13] G. S. Yovanof and S. Liu, “Statistical analysis of the DCT coefficients and their quantization error,” in *Proc. 13th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, 1996, pp. 601–605.
- [14] R. L. Joshi and T. R. Fischer, “Comparison of generalized Gaussian and Laplacian modeling in DCT image coding,” *IEEE Signal Process. Lett.*, vol. 2, no. 5, pp. 81–82, May 1995.
- [15] X. Zhao, J. Sun, S. Ma, and W. Gao, “Novel statistical modeling, analysis and implementation of rate-distortion estimation for H.264/AVC coders,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, pp. 647–660, May 2010.
- [16] Y. Altunbasak and N. Kamaci, “An analysis of the DCT coefficient distribution with the H.264 video coder,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, May 2004, pp. 177–180.
- [17] L. Chen and I. Garbacea, “Adaptive  $\lambda$  estimation in Lagrangian rate-distortion optimization for video coding,” *Proc. SPIE*, vol. 6077, p. 60772B, Jan. 2006.
- [18] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, and S. K. Mitra, “Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 182–190, Apr. 1996.

- [19] J. Choi and D. Park, "A stable feedback control of the buffer state using the controlled Lagrange multiplier method," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 546–558, Sep. 1994.
- [20] J. Zhang, X. Yi, N. Ling, and W. Shang, "Context adaptive Lagrange multiplier (CALM) for rate-distortion optimal motion estimation in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 820–828, Jun. 2010.
- [21] 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.
- [22] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.
- [23] F. Zhang and D. R. Bull, "A perception-based hybrid model for video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1017–1028, Jun. 2016.
- [24] F. Zhang and D. R. Bull, "A parametric framework for video compression using region-based texture models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1378–1392, Nov. 2011.
- [25] M. Bosch, F. Zhu, and E. J. Delp, "Segmentation-based video compression using texture and motion models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1277–1281, Nov. 2011.
- [26] P. Ndjiki-Nya, T. Hinz, and T. Wiegand, "Generic and robust video coding with texture analysis and synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 1447–1450.
- [27] P. Ndjiki-Nya, D. Doshkov, H. Kaprykowsky, F. Zhang, D. Bull, and T. Wiegand, "Perception-oriented video coding based on image analysis and completion: A review," *Signal Process., Image Commun.*, vol. 27, no. 6, pp. 579–594, 2012.
- [28] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, Apr. 2012.
- [29] C. Yeo, H. L. Tan, and Y. H. Tan, "On rate distortion optimization using SSIM," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1170–1181, Jul. 2013.
- [30] X. Wang, L. Su, Q. Huang, and C. Liu, "Visual perception based Lagrangian rate distortion optimization for video coding," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1653–1656.
- [31] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2505–2508.
- [32] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), "High Efficiency Video Coding," ISO/IEC Standard 23008-2, 2015.
- [33] F. Bossen, *Common Test Conditions and Software Reference Configurations*, document JCTVC-L1100, ITU-T, ISO/IEC, 2013.
- [34] K. Andersson *et al.*, *Non-Normative HM Encoder Improvements*, document JCTVC-W0062, ITU-T, ISO/IEC, 2016.
- [35] K. Andersson, P. Wennersten, J. Samuelsson, J. Ström, P. Hermansson, and M. Pettersson, *AHG 3 Recommended Settings for HM*, document JCTVC-X0038, ITU-T, ISO/IEC, 2016.
- [36] B. Li, L. Li, J. Zhang, J. Xue, and H. Li, *Encoding With Fixed Lagrange Multipliers*, document JCTVC-J0242, ITU-T, ISO/IEC, 2012.
- [37] R. Péteri, S. Fazekas, and M. J. Huiskes, "DynTex: A comprehensive database of dynamic textures," *Pattern Recognit. Lett.*, vol. 31, no. 12, pp. 1627–1632, 2010. [Online]. Available: <http://projects.cwi.nl/dyntex/>
- [38] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, "A video texture database for perceptual compression and quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2781–2785.
- [39] K. Sharman and K. Suehring, *Common Test Conditions*, document JCTVC-X1100, ITU-T, ISO/IEC, 2016.
- [40] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, "An adaptive QP offset determination method for HEVC," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 4220–4224.
- [41] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation. A review," *IEEE Signal Process. Mag.*, vol. 23, no. 2, pp. 28–37, Mar. 2006.
- [42] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standard—Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.
- [43] S. Winkler, "Analysis of public image and video databases for quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 616–625, Oct. 2012.
- [44] F. M. Moss, K. Wang, F. Zhang, R. Baddeley, and D. R. Bull, "On the optimal presentation duration for subjective video quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 11, pp. 1977–1987, Nov. 2016.
- [45] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, ITU-T, 2001.



**Fan Zhang** received the B.Sc. (Hons) and M.Sc. degrees from Shanghai Jiao Tong University, in 2005 and 2008, respectively, and the Ph.D. degree from the University of Bristol in 2012. He is currently a Senior Research Associate with the Visual Information Laboratory, Department of Electrical and Electronic Engineering, University of Bristol, on projects related to perceptual video compression and immersive video processing. His research interests include perceptual video compression, video quality assessment, and immersive video formats.



**David R. Bull** (M'94–SM'07–F'12) received the B.Sc. degree from the University of Exeter, Exeter, U.K., in 1980, the M.Sc. degree from The University of Manchester, Manchester, U.K., in 1983, and the Ph.D. degree from Cardiff University, Cardiff, U.K., in 1988.

He has previously been a Systems Engineer with Rolls Royce, Bristol, U.K. and a Lecturer at the University of Wales, Cardiff. He joined the University of Bristol in 1993, where he is currently the Chair of signal processing and the Director of the Bristol Vision Institute. In 2001, he co-founded a university spinoff company, ProVision Communication Technologies Ltd., specializing in wireless video technology. He has authored over 450 papers on the topics of image and video communications and analysis for wireless, internet, and broadcast applications, and has numerous patents, several of which have been exploited commercially. He has received two IEE Premium awards for his work. He is the author of three books, and has delivered numerous invited/keynote lectures and tutorials. He is a fellow of the Institution of Engineering and Technology.