



OPEN ACCESS

EDITED BY

Marco Cagnazzo,
Télécom ParisTech, France

REVIEWED BY

Luis Da Silva Cruz,
University of Coimbra, Portugal
Søren Forchhammer,
Technical University of Denmark,
Denmark

*CORRESPONDENCE

Daniel Joseph Ringis,
✉ ringisd@tcd.ie
Vibhoothi,
✉ vibhoothi@tcd.ie

RECEIVED 13 April 2023

ACCEPTED 12 May 2023

PUBLISHED 03 July 2023

CITATION

Ringis DJ,
Vibhoothi, Pitié F and Kokaram A (2023),
The disparity between optimal and
practical Lagrangian multiplier estimation
in video encoders.
Front. Sig. Proc. 3:1205104.
doi: 10.3389/frsip.2023.1205104

COPYRIGHT

© 2023 Ringis, Vibhoothi, Pitié and
Kokaram. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

The disparity between optimal and practical Lagrangian multiplier estimation in video encoders

Daniel Joseph Ringis*, Vibhoothi*, François Pitié and
Anil Kokaram

SigmaMedia Group, Department of Electronic and Electrical Engineering, Trinity College Dublin, Dublin, Ireland

With video streaming making up 80% of the global internet bandwidth, the need to deliver high-quality video at low bitrate, combined with the high complexity of modern codecs, has led to the idea of a per-clip optimisation approach in transcoding. In this paper, we revisit the Lagrangian multiplier parameter, which is at the core of rate-distortion optimisation. Currently, video encoders use prediction models to set this parameter but these models are agnostic to the video at hand. We explore the gains that could be achieved using a *per-clip* direct-search optimisation of the Lagrangian multiplier parameter. We evaluate this optimisation framework on a much larger corpus of videos than that has been attempted by previous research. Our results show that per-clip optimisation of the Lagrangian multiplier leads to BD-Rate average improvements of 1.87% for x265 across a 10 k clip corpus of modern videos, and up to 25% in a single clip. Average improvements of 0.69% are reported for libaom-av1 on a subset of 100 clips. However, we show that a per-clip, per-frame-type optimisation of λ for libaom-av1 can increase these average gains to 2.5% and up to 14.9% on a single clip. Our optimisation scheme requires about 50–250 additional encodes per-clip but we show that significant speed-up can be made using proxy videos in the optimisation. These computational gains (of up to x200) incur a slight loss to BD-Rate improvement because the optimisation is conducted at lower resolutions. Overall, this paper highlights the value of re-examining the estimation of the Lagrangian multiplier in modern codecs as there are significant gains still available without changing the tools used in the standards.

KEYWORDS

video coding, compression, optimisation, AV1, HEVC, rate-distortion optimisation

1 Introduction

User-generated video content has increased significantly in recent years (Wang et al., 2019) to the point that it now represents the majority 80% of all internet traffic (Cass, 2014) and still continues to grow. Because of the additional pressures on internet distribution and globally available bandwidth, there has always been an effort to re-engineer existing video codecs to hit better rate/quality trade-offs. Ideally, a compressed video would have zero distortion at as low a bitrate as possible, but we know that is not always practical. So decisions to achieve the optimal trade-off between bitrate and distortion need to be made. This is known as Rate Distortion Optimisation (RDO).

At the heart of this rate-distortion trade-off is the Lagrangian multiplier approach which was advocated in 1998 by Sullivan and Wiegand (1998) and Ortega and Ramchandran

(1998), and which continues to be the core mechanism inside modern video codecs. In this approach, the contained constrained optimisation problem of minimising the distortion while staying below a certain target bitrate is recast into an unconstrained minimisation of a combined cost $J = D + \lambda R$, where D is a measure of distortion, R is the bitrate and λ is the Lagrangian multiplier. The Lagrangian multiplier λ is selected to allow for different operating points in the rate-distortion curve.

Finding the value of λ that yields a target bitrate is the problem faced by the encoder. Prediction models for λ have been established as functions of the quantiser parameter QP by Sullivan and Wiegand (1998) and Ortega and Ramchandran (1998) and deployed in H.264/AVC and HEVC reference encoders. Similar models have been established for VP9 and AV1.

These models have been calibrated over video test sets but we note that these models are not tailored to a particular clip. This means that there are potentially significant gains to be made by using a per-clip optimal value of λ during encoding. Some previous works (Im and Chan, 2015; Papadopoulos et al., 2016; Yang et al., 2017; Zhang and Bull, 2019) have presented improved models to make this adaptive prediction possible. While these earlier works have shown that some gains were possible, they do not indicate what could be gained by using the optimal values of λ .

The main contribution of this paper is to propose a direct-search algorithm to optimise λ on a per-clip basis. Contrary to previous works in the academic literature that have so far only reported empirical studies on relatively small, low-resolution corpora of videos, we propose to explore the results of our method on a large video corpus which better represents today's video.

Using this direct-search for λ on a corpus of 10 K clips, we show that an average Bjøntegaard-Delta Rate (BD-Rate) improvement of 1.87% is achievable for x265 (using PSNR-Y quality metric), with about half of the corpus showing BD-Rate improvements of greater than 1% as initially detailed in Ringis et al. (2020a; b). On a subset of 100 of these videos, we show that average BD-Rate (MS-SSIM) improvements for libaom-av1 are a more modest 0.69% for libaom-av1, but performing the optimisation to a per-clip and per-frame-type basis, allows us to bring these gains to 2.47% for libaom-av1. Direct-search optimisation requires multiple measurements of the BD-Rate and, as a result, ten to fifty encodes might be required to estimate optimal values of λ using our framework. We show that the use of video proxies in the optimisation yields significant speed-up gains (x265:x21, libaom-av1:x230), whilst achieving comparable average gains (x265:0.82%, libaom-av1: 2.53%). This expands on work performed in Ringis et al. (2021) where the objective quality metric used was limited to PSNR-Y, and expands the work from (Vibhoothi et al., 2022a; Vibhoothi et al., 2022b) by using a larger amount of videos.

The rest of this paper is organised as follows. Section 2 provides a detailed review of the Lagrangian multipliers application in video encoders as well as a review of the relevant literature in per-clip Lagrangian multiplier optimisation. Section 3 details our methodology for per-clip Lagrangian multiplier optimisation, including details about our implementation. Section 4 describes each experiment undertaken and the results which are then described in Section 5. This leads to our conclusions and future work described in Section 6.

2 Background

Within a video codec, there are a number of processes and modules which have the same common goal: minimise bitrate and maximise quality. For example, in HEVC, the video coding system needs to make a decision for many parameters for each frame and each macroblock. These include Intra/Inter prediction modes, CTU/MB segmentation, and quantization step sizes. All of these decisions contribute to the shared goal of minimising distortion (i.e., maximising quality) while also attaining a low bitrate. The task of rate-distortion optimisation is to choose the parameters that achieve this goal.

2.1 Rate-distortion optimisation in video compression

2.1.1 Constrained optimisation and Lagrangian multiplier

Rate-distortion optimisation can be seen as a constrained problem, which either represents the minimisation of the bitrate, R , while keeping the distortion, D , below a set target, D_{\max} , or, alternatively, the minimisation of the distortion, D while keeping the rate below a target bitrate, R_{\max} . The optimisation is based on tuning factors including quantisation step-size, QP and macroblock selection, M and other parameters. In the following, we will consider the optimisation problem, that is:

$$\begin{aligned} & \underset{QP, M, \dots}{\text{minimize}} && D(QP, M, \dots) \\ & \text{subject to} && R(QP, M, \dots) < R_{\max}. \end{aligned} \quad (1)$$

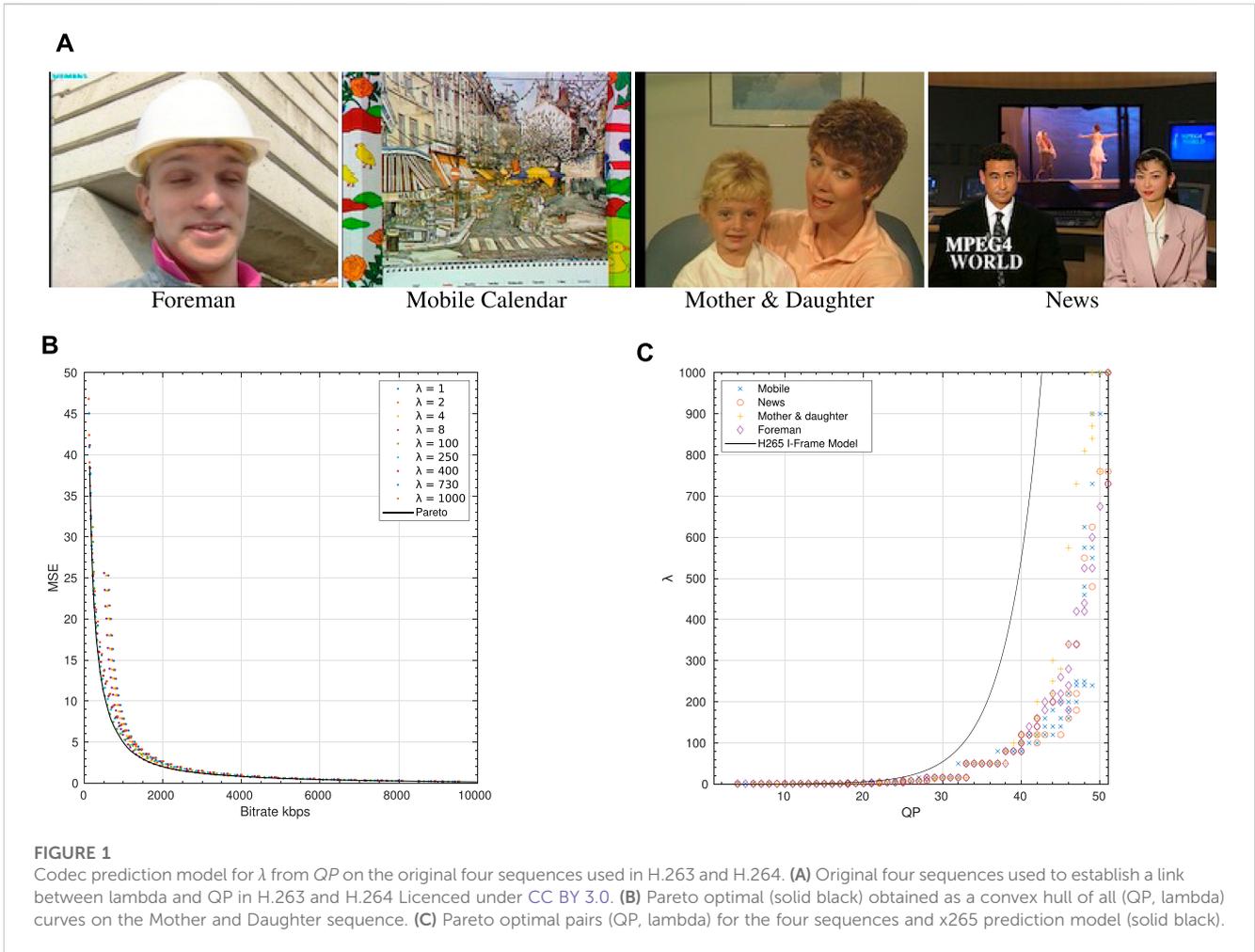
This problem can be solved using Lagrangian optimisation (Everett, 1963; Ortega and Ramchandran, 1998; Sullivan and Wiegand, 1998). In this approach the codec transforms the constrained optimisation into an unconstrained problem by introducing the cost function J as follows:

$$J = D + \lambda R. \quad (2)$$

J combines both the distortion D (for a frame or macroblock) and the rate R (the number of coded bits for that unit) through the action of the Lagrangian multiplier λ . This technique was adopted as it is effective and conceptually simple. For each value of λ , minimising J with respect to the parameters yields an optimal solution to Eq. 1 for a certain R_{\max} . Conversely for any R_{\max} , there exists a value of λ , which can be used for an unconstrained optimal solution. However, in practice, there are a number of interactions between coding decisions which make this problem less straightforward.

2.1.2 A prediction model for the Lagrangian multiplier

The first issue is how to find the optimal value of λ for a given targeted rate. The seminal works of Sullivan and Wiegand (1998); Ortega and Ramchandran (1998) laid the foundation for an experimental approach to choosing λ . The coding fidelity is principally controlled by the quantisation step QP , with a small quantiser step size leading to a high bitrate and a small amount of distortion. Considering the quantisation effect in isolation, i.e., taking QP as the sole parameter of interest, it is possible to



propose some reasonable models for the distortion and the rate as a function of QP . There is indeed a well-known high bitrate approximation (small QP approximation) in signal compression that states (see [Jayant and Noll, 1984](#)):

$$R(D) \approx a \ln\left(\frac{b}{D}\right), \tag{3}$$

where a and b are two parameters that vary depending on the video content. Combining this with the other well-known approximation of the mean squared error D for a uniform quantiser at high-bitrates ($D(Q) = \frac{QP^2}{12}$) yields a relationship between R and QP :

$$R(QP) = a \ln\left(\frac{12b}{QP^2}\right). \tag{4}$$

A remarkable property of the Lagrangian multiplier is that, at the optimum,

$$\lambda = -\frac{dD}{dR}. \tag{5}$$

This can be geometrically interpreted as λ being the negative of the slope of the optimal Pareto rate-distortion front. This can be exploited to give us a predictive model for λ :

$$\lambda = -\frac{dD}{dR} = -\frac{dD/dQP}{dR/dQP} = -\frac{QP/6}{-2a/QP} = \frac{QP^2}{12a}, \tag{6}$$

where a needs to be empirically measured.

2.1.3 Implementations of the Lagrangian multiplier prediction model in video codecs

For the H.263 Video Compression Standard, [Sullivan and Wiegand \(1998\)](#) and [Ortega and Ramchandran \(1998\)](#) experimentally determined a by encoding 100 frames from 4 sequences. They found the best fit for λ to be

$$\lambda = 0.85 \times QP^2. \tag{7}$$

An example frame from these four QCIF resolution sequences *Mother and Daughter*, *Foreman*, *Mobile and Calendar* and *News*, available on [Xiph.org](https://xiph.org) (2018), can be seen in [Figure 1A](#).

In later video codecs, H.264 and H.265, bi-directional frames (B-Frames) were introduced and the experiments to establish links between λ and QP were repeated ([Sullivan et al., 2012](#)), leading to updated relationships for each of the Intra (I), Predicted (P) and B frames. Similar efforts were led in VP9 ([Mukherjee et al., 2013](#)) and AV1 ([Chen and Murhejee, 2018](#)). In AV1, λ was estimated by optimisation over a different/

modern test video clip set, using somewhat similar prediction models:

$$\lambda = A \times q_{dc}^2, \quad (8)$$

where A is a constant depending on frame type ($3.2 \leq A \leq 4.35$) and q_{dc} is the DC quantiser.

As the encoding scenarios have evolved over the years, encoders have introduced different new frame types and used target distortion functions (e.g., VMAF over SSE), etc. To make things more complex, encoders have introduced a number of new Lagrangian multipliers for other RDO decisions that are derived from the main prediction model.

Each of these have different cost functions based on distortion, and are linked together in existing implementations of codecs. This is a practical necessity, however, we can see that codec implementations have eroded the optimality of estimation of the Lagrangian Multiplier.

2.1.4 Disparity between predictions and optimal values

To better sense how far these models deviate from optimality, we compare x265 predictions of λ with the actually measured ground-truth values of λ . For that experiment, we take the same four sequences used by Sullivan and Wiegand (1998) (Figure 1A) for h263.

We encoded each clip with x265 using I-Frames only at different QP values (1:51) for different λ values (1, 2, 4, 8, 100, 250, 400, 730 and 1,000). In Figure 1B, we report the (R , D) point cloud of all these operating points on the Mother and Daughter sequence. The Pareto front (black solid line), corresponds to the optimal $\lambda - QP$ relationship for that clip. In Figure 1C we report the (λ , QP) pairs found on the Pareto curves for the four sequences. We also plot the x265 model prediction for λ as a function of QP .

As expected, we see in Figure 1C each clip exhibits a different optimal relationship between λ and QP , and all differ from the x265 prediction model. Although the empirical relationships used in encoders perform well on average, this confirms our intuition that for any particular clip, the empirical estimate is unlikely to be optimal, and hence motivates our idea of a per-clip optimisation of λ .

2.2 Per-clip Lagrangian multiplier prediction

The general idea of per-clip parameter estimation is not new and is sometimes known as per-title, per-scene, per-shot or per-clip encoding. Netflix was the first (Aaron et al., 2015) to show that an exhaustive search of the transcoder parameter space can lead to significant gains when building the bitrate ladder for a particular clip. Those gains easily compensate for the large once-off computational cost of transcoding because that clip may be streamed millions of times across many different CDNs thus saving bandwidth and network resources in general. That idea has since been refined into a more efficient search process across shots and parameter spaces (Katsavounidis and Guo, 2018). They achieved a 13% BD-Rate improvement over the HEVC reference

codec parameters on the UltraVideo dataset (Mercat et al., 2020). Similarly, Satti et al. (2019) proposed a per-clip approach which was based on a model linking the bitrate, resolution and quality of a video. Using this model allowed for a bitrate ladder to be generated for a given clip at a lower computational cost than Katsavounidis and Guo (2018).

In the following sections, we review the limited amount of existing work on per-clip adaptation of λ (Zhang and Bull, 2019). All previous papers attempt to adjust λ away from the codec default by using a constant k such that

$$\lambda = k \times \lambda_o, \quad (9)$$

where λ_o is the default Lagrangian multiplier estimated in the video codec, and λ is the updated Lagrangian. We group these previous efforts into 3 themes, presented next.

2.2.1 Classification

One proposed method to determine k is by classifying the content of the video. In particular, Ma et al. (2016) proposed to use a Support Vector Machine to determine k from a set of image features derived from the frames. Experiments on the 37 dynamic texture videos of Ghanem and Ahuja (2010) result in improvements of up to 2 dB in PSNR and 0.05 in SSIM at the same bitrate.

Hamza et al. (2019) uses scene classification (SegNet (Kendall et al., 2015)) to classify a clip into indoor/outdoor/urban/non-urban classes. The value for k is adjusted for each macroblock, based on the class of that macroblock. They report on the Xiph.org (2018) dataset up to 6% BD-Rate improvement on intra frames. This work recognises that visual information semantics that describes the nature of the image was not considered in the experimental determination of λ for the MPEG-based codecs (H263, H264, H265).

2.2.2 Regression

Another approach proposed by Yang et al. (2017) is to determine k through a straightforward linear regression of perceptual features extracted from the video. The regression parameters are determined experimentally using a corpus from Xiph.org (2018). They report a BD-Rate improvement of up to 6.2%. Similarly to Hamza et al. (2019), they highlight that the perceptual features of the video may be of great importance in determining an optimal Lagrangian multiplier.

Zhang and Bull (2019) showed that a single feature, the ratio between the MSE of P and B frames, could give a good idea of the temporal complexity of a clip. They showed on the low-resolution (352×288) video clips of the DynTex database, that the existing Lagrangian multiplier prediction models are not ideal and that up to 7% improvement in BD-Rate could be gained.

Despite the limited scope of this experimental study, this work highlights the non-optimality of current prediction models.

2.2.3 Quantiser manipulation

Since λ models are linked to QP , another common approach is to adjust λ implicitly through the quantiser parameter, QP . This achieves a similar goal of attempting to improve the RDO of a codec, but it has a wider impact as it changes the DCT coefficients in

the compressed media as well. Taking a local, exhaustive approach, Im and Chan (2015) proposed encoding a frame multiple times in a video. Each frame was encoded using a Quantiser Parameter $QP \in (QP, QP \pm 1, QP \pm 2, QP \pm 3, QP \pm 4)$. They chose the QP which lead to the best bitrate improvement per frame. They report up to 14.15% BD-Rate improvement on a single sequence. Papadopoulos et al. (2016) also applied this idea to HEVC and propose to update QP based on the ratio of the distortion in P and B (D_P, D_B) frames of the previous Group of Pictures (GOP). They update $QP = a \times (D_P/D_B) - b$, where a, b are constants determined experimentally. They report an average BD-Rate improvement of 1.07% on the DynTex dataset, with up to 3% BD-Rate improvement achieved for a single sequence.

2.3 Remarks

Working practical codec implementations have deviated from the optimality of the theory of Lagrangian multipliers. That is because the proposed prediction models make a number of assumptions (e.g., QP is the only decision parameter, QP is assumed to be small, etc.), and also because some considerations simply break the theory (e.g., the definition of rate and distortion costs differ in the motion estimation module). This was inevitable because optimal estimation is just too expensive to achieve *per-clip*. To make matters worse, historically, experiments were performed on small corpus sizes and on content which does not adequately represent modern media. Hence, it is likely that a better λ exists for an individual video clip which improves the BD-Rate compared to the current defaults used in existing codec implementations.

Our first observation is that all previous works confirm this hypothesis, and they also show that better adaptive determinations of λ are possible and can yield an increase in rate-distortion performance.

Secondly, all previous works propose new models to make this adaptive prediction possible. These models are also themselves necessarily sub-optimal, as only a brute-force exploration of k from Eq. 9 would lead to the optimum.

Finally, the scope of these previous experimental studies is limited as they have been performed on small number of mostly low-resolution videos that bear little resemblance to modern video content, e.g., as found on popular streaming platforms.

3 A framework for a per-clip direct-search optimisation of the Lagrangian multiplier

To explore just how much more BD-Rate gains are to be gained using an adaptive approach, we propose in this paper to find the best possible λ per-clip, using a direct-search optimisation of λ . Similarly to the rest of the literature, we adjust λ away from the codec default estimation λ_o through the use of a constant multiplying factor, k , across the clip/sequence, as $\lambda = k \times \lambda_o$. Our objective is to apply a direct-search optimisation of k . Our experiments target x265 (HEVC) and libaom-av1 (AV1), and contrary to previous studies, our experiments are performed on a modern corpus of videos, built on the large YouTube UGC dataset (Wang et al., 2019). Lastly, motivated by the fact that the prediction model for λ is already

dependent on the frame type in the existing libaom-av1 implementation, we study the effect of isolating the optimisation of λ for different frame types in libaom-av1.

3.1 BD-rate objective function

The objective function for our direct-search optimisation is directly chosen to be the BD-Rate (Bjontegaard, 2001; Tourapis et al., 2017).

In the standards, the BD-Rate is computed in the log domain. Defining r as $\log(R)$, it can be implemented as:

$$\Delta_{\text{BDR}} = \exp(\mathbb{E}[r_2 - r_1]) - 1, \quad (10)$$

where $\mathbb{E}[r_2 - r_1]$ is defined as

$$\mathbb{E}[r_2 - r_1] = \frac{1}{Q_2 - Q_1} \int_{Q_1}^{Q_2} (r_k(Q) - r_1(Q)) dQ, \quad (11)$$

and the integral is evaluated over the quality range $[Q_1, Q_2]$. $r_1(Q)$, $r_k(Q)$ are the RQ-Curves corresponding to λ_o and $\lambda = k\lambda_o$ respectively. Optimisation for multiple frame types can be simply achieved by employing multiple values of k associated with each frame type (e.g., $\lambda_I = k_I\lambda_o$ and $\lambda_P = k_P\lambda_o$ if we jointly optimise for I and P frames respectively). Each RD operating point is generated using the same rate control mode (e.g., CRF, CQP) within a range that matches typical streaming media use cases. Then a curve fit, as recommended in (Bjontegaard, 2001; Tourapis et al., 2017), is used for evaluating the integral over the entire $[Q_1, Q_2]$ range.

Note that for every evaluation of the BD-Rate, a number of encodes are required. This is because each evaluation of the BD-Rate requires the generation of a full rate-distortion curve. In our experiments, we chose 5 RD points as a reasonable trade-off between computational complexity and precision of the estimated RQ-Curve.

3.2 Optimisation methods

A number of off-the-shelf optimisation techniques are available to us to minimise the BD-Rate objective function. For the one-dimensional search of a single k value, traditional direct-search methods such as Brent's method are well-established for finding the local minimum of such a unimodal function. Other minimization solvers could be used in place of Brent's Method. However, this was selected for its quick off-the-shelf use (Press et al., 1992) for one-dimensional search.

To jointly optimise for two or more values of k (see section 4.3), we propose to use Powell search technique (Press et al., 1992). This is a conjugate direction method performing a sequential one-dimensional search along each direction in a multiple multiple direction set. We experimented with other optimisation methods such as downhill Simplex, or Conjugate Gradient, but the cost function surface is too flat, and Powell was consistently better.

3.3 Implementation

Encoding Settings. For each codec, the codebase was modified to take k as an argument. Hence this alters the rate-distortion

constraint from Eq. 9 used in RD control throughout the codec. The value of k is used at a top level to alter the default Lagrangian multiplier ($\lambda = k \times \lambda_0$). The main encoder decisions impacted by this change include frame partitioning, motion vector selection, and mode decision. Each of these decisions tries to minimise the Rate-Distortion cost function ($J = D + \lambda R$), and a change of λ will change that RD balance. To investigate a frame-level optimisation in AV1, we have modified the codebase to allow for k to be changed at a frame-type level.

For H.265, we used the x265¹ implementation. For AV1, we used the libaom-av1² implementation which is the research reference codebase. The open-source nature of these implementations was the main reason for selecting them. As there was no previous study with AV1, we chose to use the reference implementation for reproducibility. As our experiments used a corpus of nearly ten thousand videos, we selected the open source codec x265 based on its encoding speed compared to its reference model counterpart HEVC-HM.

RD-Curve Generation. We use 5 RD points for both H.265 and AV1 based on Common-testing configurations of both codecs (Bossen and Jan 2013; Zhao et al., 2021). For H.265, we use $QP \in \{22, 27, 32, 37, 42\}$ and for AV1, $QP \in \{27, 39, 49, 59, 63\}$. For the encoding configuration in x265, we deployed constant rate factor (CRF) mode and in libaom-av1, we deployed Random-Access configuration. It is important to note that this work does not aim to compare HEVC and AV1 but to improve both codecs relative to their default behaviour.

Quality/Distortion Metrics. In this work, we deployed two objective quality/distortion metrics D for evaluation. For the larger UGC-10 k dataset (Section 3.4.1 below), we report PSNR-Y as it is the most common objective quality metric in academic literature and has the lowest compute cycles for measuring quality points for around 10,000 videos with multiple RD-points per optimisation study. Despite being the most commonly reported metric, PSNR-Y has been shown to not correlate strongly with perceived quality Wang et al. (2003). For the smaller dataset of 100 videos (UGC-100, see below), we deployed MS-SSIM as the objective quality metric. MS-SSIM captures structural similarity based on human perception which is also not computationally expensive compared to other perceptual metrics.

3.4 Dataset

Previous works (Ma et al., 2016; Zhang and Bull, 2019) only used a small corpus size of approximately 40 clips, with up to 300 frames per-clip. The type of content used in these previous corpora is also not necessarily a good representation of modern material. We propose therefore to build on the YouTube UGC dataset (Wang et al., 2019). This User Generated Content (UGC) dataset is a sample of videos uploaded to YouTube, and contains about 1,500 videos across 12 categories (Figure 2A), and video resolutions ranging from 360p

to 4 k. From this UGC dataset, we formed an expanded version (UGC-10 k), which we will use for large-scale analysis with HEVC, and another smaller subset of 100 clips (UGC-100), to allow analysis with AV1.

3.4.1 UGC-10k

Our UGC-10k dataset consists of $9,746 \times 150$ -frame clips with varying framerates representing DASH segments. In addition to the UGC videos, we use clips from other publicly available datasets, these include the Netflix dataset (Chimera and El Fuente) (Netflix, 2015), DynTex dataset (Ghanem and Ahuja, 2010), MCL (Lin et al., 2015) and Derfs dataset (Xiph.org, 2018). Spatial resolutions from 144p to 1080p are considered. This represents more than a 100-fold increase in the amount of data used for experiments as compared to previous works.

The large corpus allows for a better representation of the application of encoding for internet video use cases. In particular, datasets like the YouTube-UGC (Wang et al., 2019) and the Netflix (Netflix, 2015) dataset directly represent the content of two of the most popular video streaming services today. User Generated Content video was not used in prior research in the area of rate-distortion optimisation. With the release of this dataset, our work highlights the limitations of the existing codec implementations on the most commonly uploaded and viewed style of video.

3.4.2 UGC-100

From the clips of the YouTube-UGC dataset, we curated a subset of 100 clips across different UGC categories with different resolutions ranging from 360p to 2160p. The videos are sampled from the 12 UGC categories. We also included 3 additional categories: HDR (2), VerticalVideo (7) and VR (4) from Wang et al. (2019). The sequences used are 130 frames instead of 150 frames.

Figure 2B shows the Spatial and Temporal Energy computed using the Video Complexity Analyzer software (Menon et al., 2022). VCA is a low-complexity spatial and temporal information extractor which is suitable for large-scale dataset analysis compared to conventional Spatial Information (SI) and Temporal information (TI) from ITU-P.910 (ITU-T RECOMMENDATION, 2022). The SI feature is derived from the energy of the DCT coefficients while the TI is derived from the frame differences.

4 Results

4.1 Per-clip direct-search optimisation

4.1.1 UGC-10k (x265)

Our first experiment examined BD-Rate (PSNR-Y) optimisation using Brent's method for x265 across the UGC-10 k dataset. The scatter plot of all the resulting BD-Rate improvements for each clip is reported in Figure 3. The videos are grouped into their respective categories. The dashed red line represents the average BD-Rate improvement for that category. In these plots, the higher the improvement, the better the performance. We see that there are a few outliers which have excellent improvement. The corpus shows on average 1%–2% improvement.

In order to compare to previous literature, we select nine clips which had reported results in Zhang and Bull (2015). These clips can be classified as live action video Akiyo, News, Silent, Bus, Tempete,

¹ Version: 3.0 + 28-gbc05b8a91.

² Version: 3.2.0–287164d.

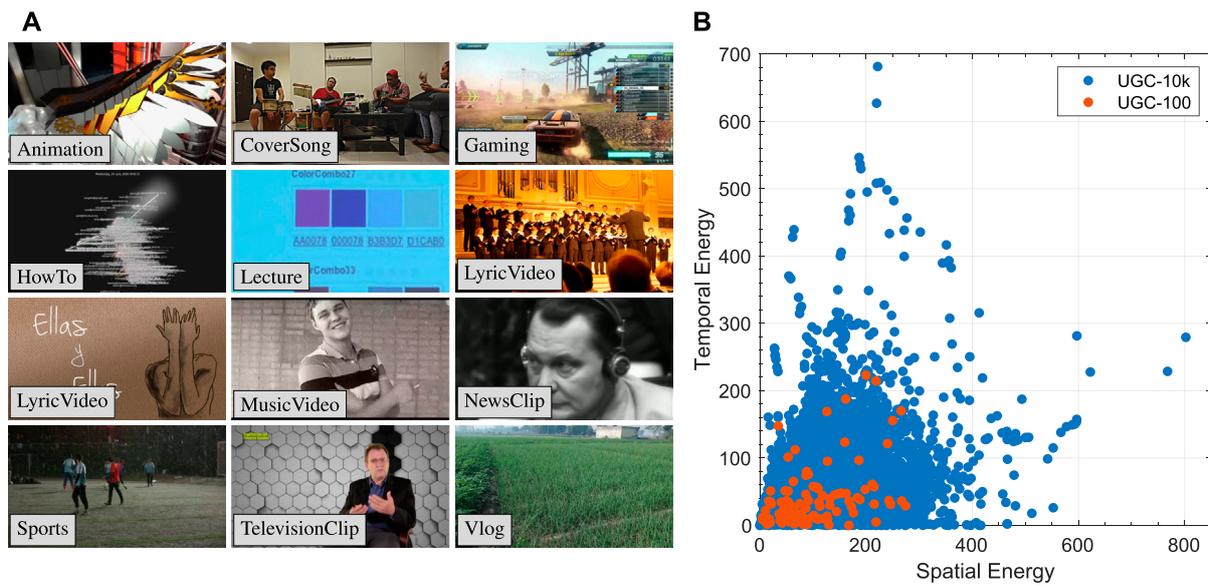


FIGURE 2

Composition of our UGC-10 k (10 k videos) and UGC-100 (100 videos) datasets. **(A)** Example frames from each of the twelve categories used in the YouTube UGC dataset. Animation_720P-6372 is an audio-removed excerpt from "Alyssa Mitchell Demo Reel 2016" by "Alyssa Mitchell" licensed under CC BY 4.0, CoverSong_720P-449f is an audio-removed excerpt from "Radioactive Cover by The Gr33nPill (Imagine Dragons Original)" by "The Gr33nPill" licensed under CC BY 4.0, Gaming_720P-6403 is an audio-removed excerpt from "NFS MW—Speedlist Gameplay 4—PC Online" by "Dennis González" licensed under CC BY 4.0, HowTo_720P-6791 is an audio-removed excerpt from "Google Chrome Source Code Progression" by "Ben L." licensed under CC BY 4.0, Lecture_360P-6656 is an audio-removed excerpt from "Lecture 8: Website Development—CSCI E-1 2011—Harvard Extension School" by "Computer Science E-1" licensed under CC BY 4.0, LiveMusic_720P-2620 is an audio-removed excerpt from "Triumph Dresdner Kreuzchor in St. Petersburg (Russia)—The final part of the concert" by "PITERINFO © ANDREY KIRILLOV" licensed under CC BY 4.0, LyricVideo_720P-4253 is an audio-removed excerpt from "Silvia Tomàs Trio—Ellas y ellos" by "Silvia Tomàs Trio" licensed under CC BY 4.0, MusicVideo_720P-7501 is an audio-removed excerpt from "RONNIE WILDHEART "HONESTLY" (OFFICIAL MUSIC VIDEO)" by "RONNIEWILDHEART" licensed under CC BY 4.0, NewsClip_720P-35d9 is an audio-removed excerpt from "День города - новости Рязани 22.01.16" by "Телекомпания Горька" licensed under CC BY 4.0, Sports_720P-6bb7 is an audio-removed excerpt from "12 тур ЛФЛ СЛ 2017/18, сезон 2:1 СЛАВИЯ 2:1 СКА (весь матч)" by "ЛФК СЕВЕР" licensed under CC BY 4.0. **(B)** Spatial and Temporal Energy distribution for UGC-10k (blue), UGC-100 (red).

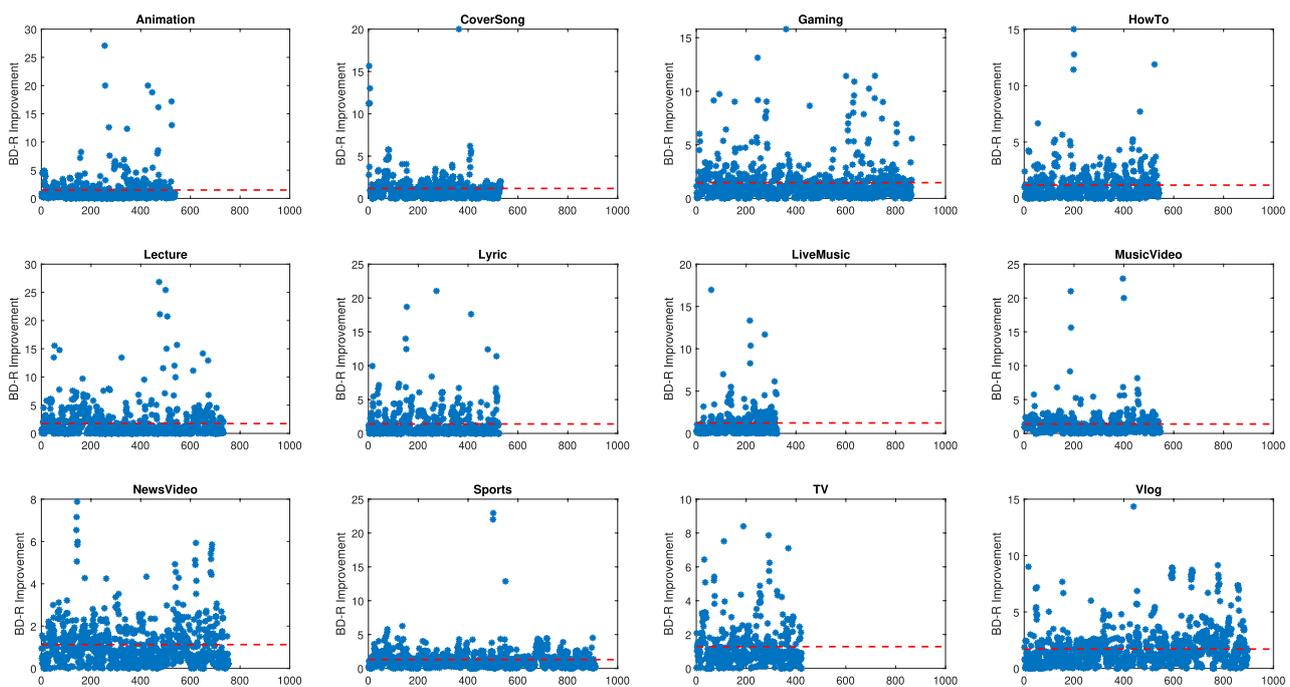


FIGURE 3

BD-Rate improvement for each clip in the YouTube-UGC corpus (higher is better). The dashed red line represents the average BD-Rate improvement for that category of video.

TABLE 1 Selected Clips showing BD-Rate improvement using our direct optimisation (BD-Rate(k_{opt})) as well as reported results from another adaptive Lagrangian Multiplier(BD-Rate(Zhang and Bull, 2019)) We see that our system is better applied to natural content of live action video Akiyo, News, Silent, Bus, Tempete, Soccer available from Xiph.org (2018) with larger BD-Rate improvement in all six clips. Unfortunately it is not as successful on the DynTex sequences Shadow, Shower, Wheat from Ghanem and Ahuja (2010) as the BD-Rate improvement reported in these three clips is larger. We believe that this is due to the nature of the DynTex clips being visually complex sequences. It is important to note that Zhang and Bull (2015) was done on HEVC-HM and our work was done on x265 which is not a direct comparison.

Sequences	Avg. BD-Rate(k_{opt}) %	Avg. BD-Rate(k_{zB}) %
	Ours, x265	Zhang and Bull (2015), HEVC
Akiyo, News, Silent	12.8	7.9
Bus, Tempete, Soccer	0.9	0.5
Shadow, Shower, Wheat	0.4	4.8

TABLE 2 BD-Rate gains for PSNR from each category of the YouTube-UGC corpus using direct-search optimisation in x265 and the overall gains found in our UGC-10 k dataset.

YouTube-UGC category	BD-Rate(%) PSNR		Clips (%) with gains greater than	
	Average	Best	1%	5%
Animation	1.48	22.07	41	4.7
CoverSong	1.17	20.00	41	2.0
Gaming	1.46	15.79	50	4.7
HowTo	1.17	15.01	40	1.8
Lecture	1.73	21.84	43	7.0
LiveMusic	1.23	16.96	37	2.5
LyricVideo	1.39	21.03	35	5.7
MusicVideo	1.38	23.86	49	3.1
NewsVideo	1.12	7.87	43	1.5
Sports	1.29	22.93	50	0.6
TV	1.28	8.39	47	2.6
Vlog	1.71	14.34	60	5.0
Overall	1.87	23.86	45	2.0

Soccer available from Xiph.org (2018) and DynTex sequences Shadow, Shower, Wheat from Ghanem and Ahuja (2010). We see in Table 1 that our system resulted in better BD-Rate improvement in the live action sequences up to 4.9% better than previous work. However, it is important to note that their work was on the HEVC-HM codec whereas ours is on x265. However, the improvement seen in the highly textured Dyntex clips was worse (0.4% compared to 4.8%) than the results reported in Zhang and Bull (2019). This is probably because our system performs better on more natural scenes.

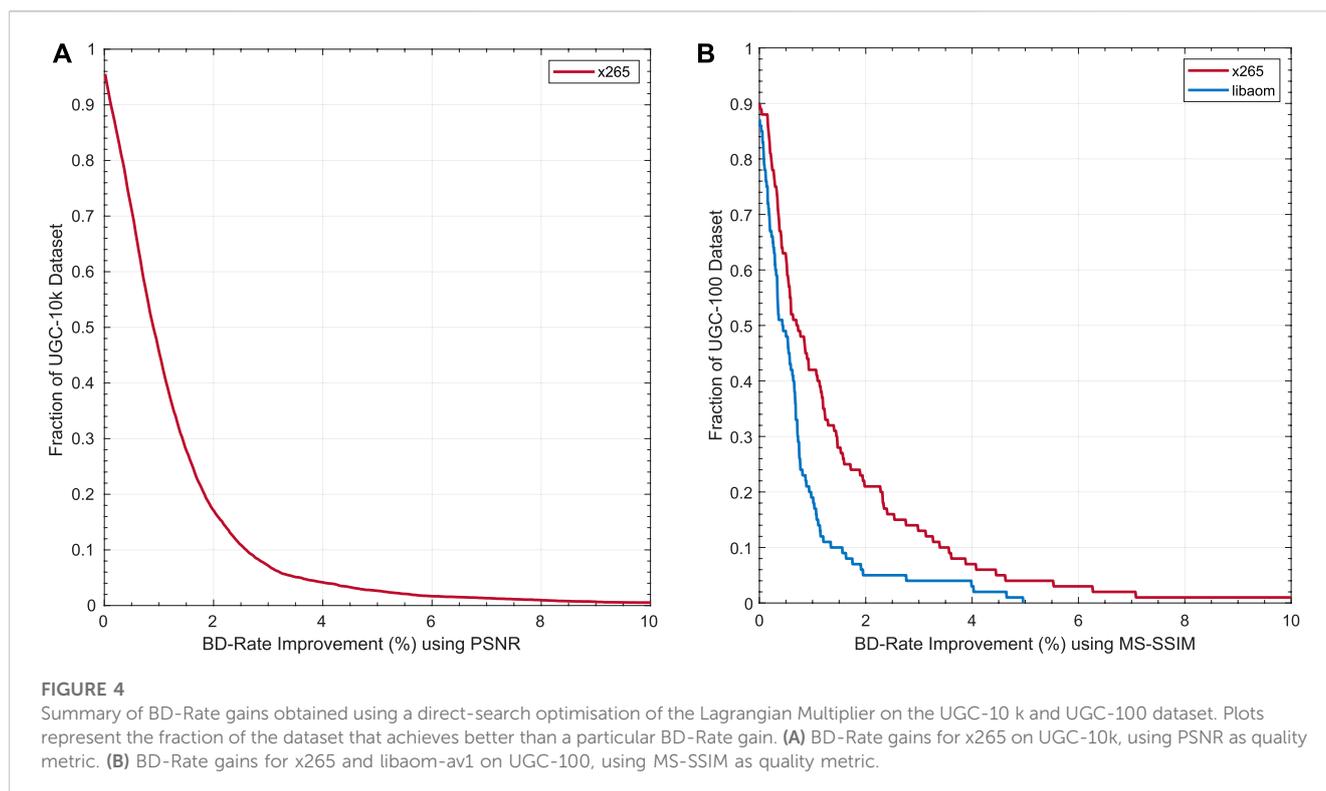
While it is useful to see the improvement of each individual clip in the corpus, it is better to quantify the impact of these experiments with a summary of the gains. For each category, we can see the best and average BD-Rate (%) improvements in Table 2. Also shown are the percentage of clips which have at least 1% and 5% improvement. In order to best represent the improvements made across our entire corpus, Figure 4A shows

the fraction of the clips encoded with our system that yield a BD-Rate improvement of at least X%. Interestingly the Animation, Lecture and Vlog categories provided the best improvements. The majority of frames in these content types exhibit low temporal complexity. That may explain why these categories gain the most from optimisation.

The results validate the usefulness of direct optimisation. The best BD-Rate improvement after optimisation is 23.86%. About 47% of our corpus shows a BD-Rate improvement greater than 1%. Of course, there are still a number of video clips where the original Lagrangian multiplier was the best (8% for x265).

4.1.2 UGC-100 (x265, libaom-av1)

A similar experiment was then run on the UGC-100 dataset for both libaom-av1 and x265. We optimised here for the MS-SSIM objective quality metric. Results summarised in Figure 4B shows that approximately 90% of the clips for either codec show some



improvement after an adjustment to the Lagrangian multiplier is applied. We achieve an average BD-Rate improvement of 1.35% in x265, with our best performers showing gains of 10.64%. We see that the results of our subset, UGC-100, provide similar results to the gains found on the larger dataset using x265. For libaom-av1, we observe more modest gains, with average BD-Rate gains of 0.69% (up to 4.96% on one clip), with 19% of the clips showing gains that are greater than 1%.

4.2 Revisiting the default prediction model in x265

We next revisit the default prediction model by estimating the best-performing value of k across the corpus and applying that to all videos. The rationale is twofold. First, w.r.t. Whether the current scaling used in the x265 prediction model is also optimal with UGC. Second, we want to check whether the observed per-clip gains could be more simply obtained by a better scaling of the default model.

To find the best value of k across the dataset, we evaluate k for each video in the range $k = 0:0.001:3$. To reduce the computational cost, this estimation was done on the UGC-10 k corpus downsampled to 144p. The graph of the average BD-Rate improvement at 144p for a given k is reported in Figure 5A. Positive improvements are observed in the 0.6:1 range, with the best average improvement at $k = 0.782$. Incidentally, $k = 0.782$ also corresponds to the most frequent optimal value of k values found with direct optimisation on the corpus videos.

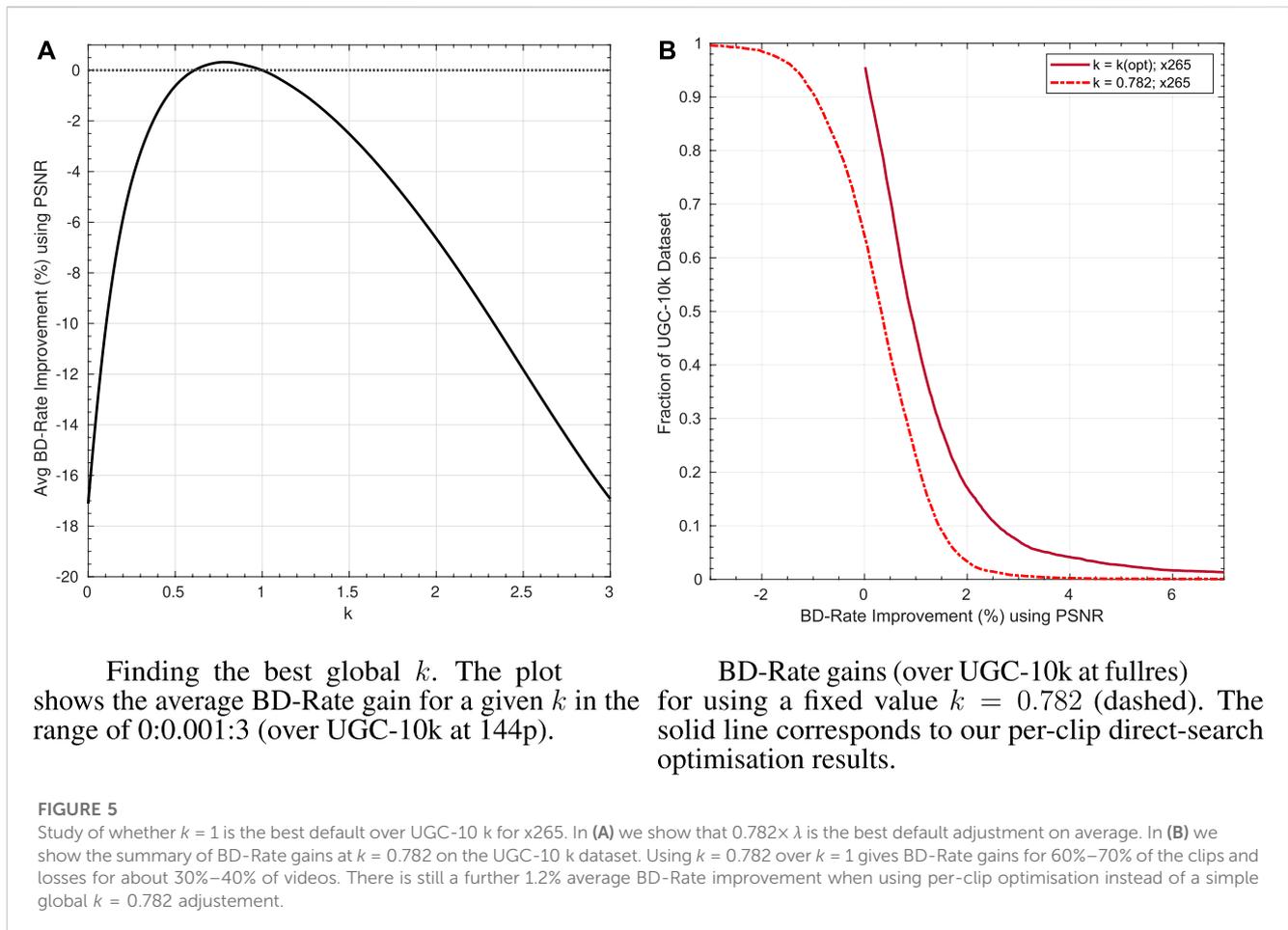
BD-Rate results for $k = 0.782$ over the actual UGC-10 k (i.e., at full resolution) are reported in Figure 5B. We observe that, for

about two-thirds of the clips, $k = 0.782$ performs better than the default $k = 1$, and for about one-third of the videos, the default settings are better. The average BD-Rate gains for $k = 0.782$ are 0.63% across our corpus with 20% of the corpus showing a BD-Rate improvement of 1% or more. This indicates that the default Lagrangian Multiplier for x265 may not be the best. Hence for x265, we would recommend adjusting the Lagrangian Multiplier to $0.782 \times$ its current value. Theoretically a better adjustment could be found by optimising k at full resolution instead of optimising at 144p proxy resolution. Our results in Section 4.4 however suggest that optimising at 144p or full resolution yields, in practice, comparable results.

We also note that per clip direct-search optimisation (i.e., $k = k_{opt}$) still yields much larger gains overall compared to applying a global optimisation to the Lagrangian multiplier. Note that the optimisation results always yield positive gains, because we always check the BD-Rate against the default.

4.3 Frame-level optimisation in AV1

As observed earlier, modern video encoders have introduced λ adjustments based on the frame type, thus we study a finer optimisation of λ at a frame-type level and we explore this idea on libaom-av1. AV1 proposes 7 reference frame types (Liu et al., 2017). We can classify them as Intra-coded frame types and Inter-coded frame types where the intra-frame is known as KEYFRAMES (KF) corresponds to the usual Intra coded reference. For inter-coding, GOLDEN_FRAME (GF) is a frame coded with higher quality and, ARF_FRAME is an alternate reference system which is used in prediction but does not



Finding the best global k . The plot shows the average BD-Rate gain for a given k in the range of 0:0.001:3 (over UGC-10k at 144p).

BD-Rate gains (over UGC-10k at fullres) for using a fixed value $k = 0.782$ (dashed). The solid line corresponds to our per-clip direct-search optimisation results.

appear in the display, it is also known as an Invisible frame. The other 4 frame types are less frequently used and will not be considered in this study. Keeping this in mind, we identified 5 different tuning mechanisms of λ with a multiplier k at different levels. They are, *i*) the previous method of using a single k for all the video frames, *ii*) tuning of the KF frames only, *iii*) tuning of both GF/ARF with a single k , *iv*) tuning of KF/GF/ARF with a single k , *v*) tuning of KF with k_1 and GF/ARF with k_2 . In all these groupings, except for (*i*), all the other frame types are encoded using the default settings.

We evaluated these methods on UGC-100 using Brent's technique of (i)–(iv) and Powell's method for (v). Table 3 reports the overall BD-Rate (%) improvements in terms of MS-SSIM. The *all-frames* tuning method (i) corresponds to our previous results for libaom-av1. The best single k tuning grouping seems to be tuning for GF/ARF together (iii), as this can boost the average gains to 1.39%, with 40 clips achieving greater than 1% gains, and 7 clips obtained greater than 5% improvement.

Further significant improvements are obtained in (v) with the joint optimisation of two k multipliers. We obtain average BD-Rate gains of 2.47%, with more than 68 and 12 clips with greater than 1% and 5% improvements respectively. In all these cases, we also achieved 5%–10% bitrate savings on different operating points with negligible loss in objective metrics of 0.07 dB.

Figure 6 shows the distribution in terms of the fraction of the dataset achieving a certain level of BD-Rate improvement. It is

evident that optimising λ for specific frame types can improve the overall BD-Rate(%) gains in libaom-av1.

4.4 Proxy optimisation

As computational complexity is of key importance given the amount of video data being processed, we explore the use of proxy systems for estimating λ at lower cost in terms of CPU cycles. The idea of optimisation on proxy videos is not new. In particular, Shen and Kuo (2018) showed that it is possible to obtain good parameter estimates, including for λ , from downsampled videos. Similarly (Wu et al., 2020; Wu et al., 2021), demonstrated that it is possible to extract features from a fast encoder proxy pass to regress the encoding settings of the higher-complexity pass.

Motivated by these studies, we propose for x265 to estimate k at a lower resolution. Videos that are less than 720p in resolution are downsampled to 144p, and higher-resolution videos are downsampled by half. For libaom-av1, similar low-resolution proxies did not yield good performance but we found that a good proxy is to estimate k at a faster speed preset (Speed-6) (cpu-used = 6) instead of the default Speed-2. Generally, when the encoder is at a faster speed-preset, a number of coding tools and encoding decisions are skipped. Specifically, the motion estimation precision is reduced, the number of inter and intra-predictions is decreased, partition pruning and early-exit conditions are more

TABLE 3 Results for different frame-level tuning modes. The best single k tuning grouping seems to be tuning for GF/ARF together (iii), as this can boost the average gains to 1.39%, with 40 clips achieving greater than 1% gains, and 7 clips obtained greater than 5% improvement. Further significant improvements are obtained in (v) with the joint optimisation of two k multipliers. We obtain average BD-Rate gains of 2.47%, with more than 68 and 12 clips with greater than 1% and 5% improvements respectively.

Frame-type tuning groups	k	BD-Rate(%) MS-SSIM		Clips (%) with gains greater than		Iterations	QP39 bitrate savings (%)	QP39 MS-SSIM loss (dB)
		Average	Best	1%	5%			
i) k : All-frames	1.34	0.69	4.96	19	0	11.71	7.60	0.36
ii) k : Key-frames (KF)	2.88	0.82	4.79	27	0	12.24	1.37	0.03
iii) k : GF, ARF	1.85	1.39	11.09	40	7	12.25	2.26	0.08
iv) k : KF, GF, ARF	1.65	1.30	7.74	40	5	11.35	1.87	0.11
v) k_1 : KF; k_2 : GF, ARF	(9.78, 1.74)	2.47	14.89	68	12	47.79	3.44	0.07

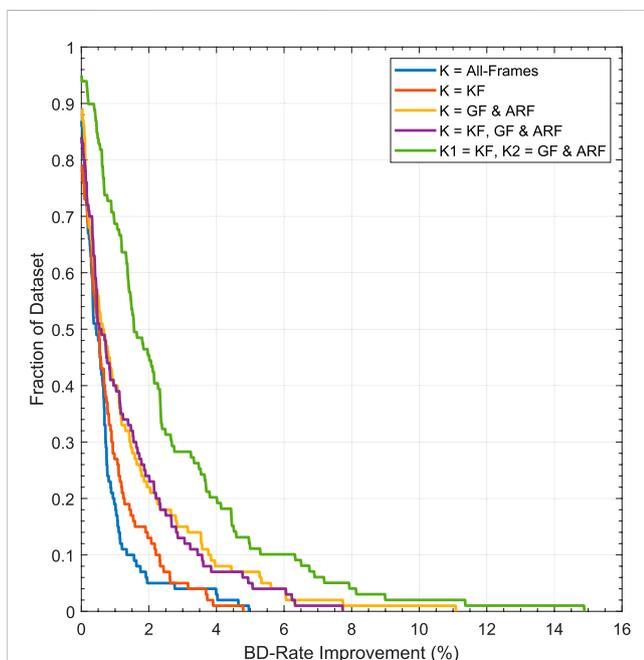


FIGURE 6

BD-Rate (%) improvement plot showing fraction of dataset achieving certain per cent of BD-Rate(%) of MS-SSIM improvement for AV1 (libaom-av1). Curves towards the top right represent a better performing system.

aggressive, and a reduced set of in-loop filters is employed. The values of k estimated on the data are then deployed at the original resolution/preset.

Proxy prediction results on the UGC-100 dataset are summarised in Table 4. The proposed estimation of k on proxy videos still leads to BD-Rate gains on a majority of video clips. In Figure 7 we see an average BD-Rate improvement across the corpus of 0.82% for our x265 proxies (vs. 1.35% without proxy) and 2.50% for our libaom-av1 proxies (vs. 2.47% without proxy). The computational complexity is greatly reduced, with a $\times 22$ speedup for x265, and a $\times 230$ speed-up for

libaom-av1. Note that the reported iteration times in Table 4 include the encoding of 10 RD points (5 for default encoder settings, 5 for encoding with a k from optimiser). Thus for our x265 proxy setup, we achieve 60% of the possible BD-Rate gains with a $\times 22$ speedup, and for our libaom-av1 proxy setup, we obtain 100% of possible BD-Rate gains at a $\times 230$ speedup.

On average, it takes 12 iterations for x265, and 40 iterations for the libaom-av1 frame-type grouping (v), so the overall overhead for the λ optimisation framework amounts to about $25.29/5817.10 \times 250 \approx 1$ one additional encode for libaom-av1 and $0.15/3.25 \times 50 \approx 2.3$ encodes for x265. The cost of downsampling the videos is not accounted for here as it would be prepared as part of the streaming pipeline in most content delivery systems.

We note that the faster speed preset proxy in libaom-av1 is a better proxy than the low-resolution proxy in x265. A possible explanation for this disparity is that the nature of the RDO problem does not change too much between the two speed presets (e.g., the choice of motion estimator should not impact the rate-distortion trade-off). Conversely, in the low-resolution proxy, the input data statistics change significantly (e.g., maximum block and transform sizes which can be used for coding a frame are different when we downsample the video).

5 Discussion

The first major takeaway is that there are significant gains available from directly optimising λ on a per-clip basis. With almost 25% BD-Rate improvement found on some clips and average BD-Rate improvements around 2%, there is a lot of potential in this kind of approach, especially considering that, as uploaded videos are streamed in the thousands to millions of views range, a bitrate savings of as small as 0.1% can have an impact. Gains can vary from video to video and codec to codec, but our experiments give us a better indication of the upper bound of what could be gained with a better λ prediction.

The proposed framework requires a number of additional encodes, but we show that estimating λ with proxy videos seems

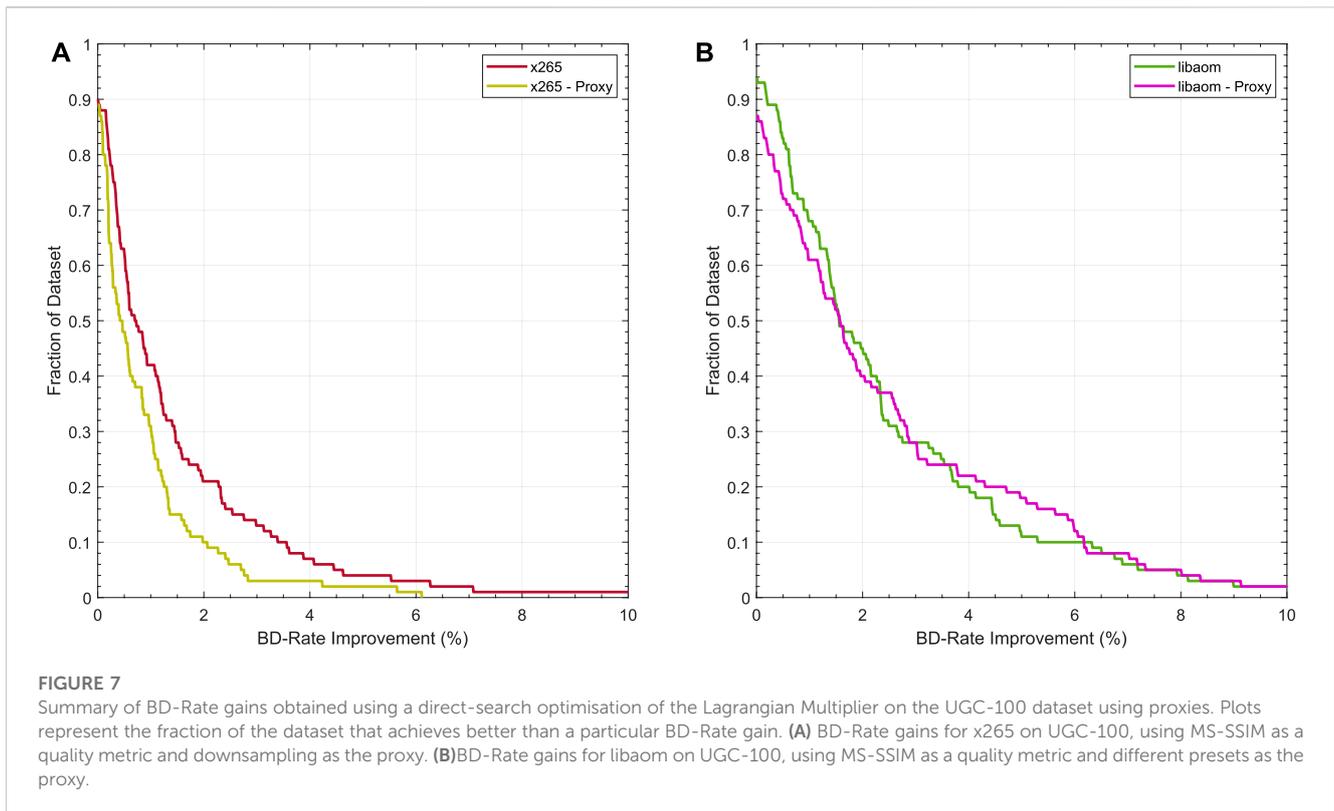


TABLE 4 BD-Rate (%) measured in terms of MS-SSIM between default and proxy-method for our dataset. Encoding time per each optimisation cycle (mins) is reported showcasing different methods can achieve substantial speedup for a fractional loss.

Codec	Method	BD-Rate(%) MS-SSIM		Clips (%) with gains greater than		Avg. Encoding time per-iteration (mins)
		Average	Best	1% gains	5% gains	
x265	Direct	1.35	10.64	42	4	3.25
x265	Proxy	0.82	6.11	30	2	0.15
Libaom-av1	Direct	2.47	14.89	65	11	5817.10
Libaom-av1	Proxy	2.50	17.41	61	17	25.29

to yield satisfying results, at 60%–100% of the potential gains, and helps reduce the additional cost. It should be possible to reduce computational costs even further. One way would be to combine the different proxy methods, e.g., using 144p resolution videos with faster codec settings. That depends on a good trade-off between the speed and the reliability of these estimates. Another route would be to model the RD-curves so as to reduce the number of operating points required to get a sufficiently accurate estimate of the BD-Rate.

Another takeaway is that a systematic optimisation over a large, representative, dataset, enables us to explore the behaviour of current codecs and can unveil some potential issues in their current implementations. For instance, we observe that x265’s default implementation should probably be $0.782 \times \lambda$. Note that this is for the open-source implementation of HEVC, this may not hold true for the reference model HEVC-HM. Another observation is that there is a lot of room for improvement when targeting a frame-level optimisation

in AV1. This may indicate that the current λ -prediction models in libaom-av1 are a bit off and that some gains could be achieved by a simple re-calibration of the current formulas. It is our recommendation that these frame-type λ formulas in the libaom-AV1 implementation should be revisited.

The exploration of the dataset results also allows us to find potentially interesting positive outliers. For instance, videos with low temporal energy, such as those found in some of the UGC animation clips, lecture slideshows and vlogs, seem to have the most potential gains for better λ prediction. This information could be used to determine which clips stand to gain the most from adjustments to λ and be prioritised in content delivery systems using our approach.

Lastly, the idea of parameter optimisation brings the focus to the choice of the objective quality/distortion metric. In this paper, we are minimising the BD-Rate with respect to PSNR-Y and MS-

SSIM. It is however likely that, in some situations, optimising for one particular quality metric will come at some cost for other metrics, possibly in a null-sum game scenario, where gains are only obtained at the expense of losses for other metrics. Future work should therefore consider complementing our analysis with a subjective study, so as to evaluate how the measured objective performance gains are actually perceived by viewers (Vibhoothi et al., 2023).

6 Conclusion

In this work, we introduced the idea of a per-clip and per-frame-type optimisation of the Lagrangian Multiplier parameter used in video compression. Whereas previous works proposed prediction models to better estimate λ , our direct-search optimisation approach allows us to establish an upper bound for the gains that can be obtained with better per-clip λ values. Also, instead of working on a handful of videos as has been done in the past, our work includes experiments on a large corpus of 10 k videos.

Our results show that BD-Rate (PSNR) average improvements of about 1.87% for x265 across a 10 k clip corpus of modern videos, and up to 25% in a single clip. For libaom-av1, we show that optimising λ on a per-frame-type basis, improves our average BD-Rate gains (MS-SSIM) from 0.69% to 2.5% on a subset of 100 clips. We also show that these estimations of λ can effectively be done on proxy videos, at a significantly reduced cost of about less than an additional encode.

Beyond the raw performance gains, this paper highlights that there is much to be learned from studying the disparity between the optimal settings and the practical predictions made by video encoders.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://media.withyoutube.com/>.

Author contributions

AK and FP contributed to the conception and design of the research. DR organised the UGC-10k dataset and implemented the experiments for x265. Vi established the UGC-100 dataset and conducted and implemented the results for libaom-av1. FP wrote the first draft of

the manuscript. DR and Vi wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was partly funded by the Disruptive Technology Innovation Fund, Enterprise Ireland, Grant No. DT-2019-0068, the ADAPT-SFI Research Center, Ireland with Grant ID 13/RC/2106_P2. YouTube and Google Faculty Awards, and the Ussher Research Studentship from Trinity College Dublin. The authors declare that this study received funding from Google. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

Thanks to Sigmedia.tv, AOMedia, YouTube Media and Algorithms Team, and other Open-Source members for helping and supporting Research and Development.

Conflict of interest

The author FP declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aaron, A., Li, Z., Manohara, M., De Cock, J., and Ronca, D. (2015). Per-title encode optimization. *Netflix Techblog*. Available at: <https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2>.
- Bjontegaard, G. (2001). Calculation of average PSNR differences between RD curves; VCEG-M33. *ITU-T SG16/Q6*.
- Bossen, F. (2013). *Common test conditions and software reference configuration*. JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 12th Meeting; Doc: JCTVC-L1100.
- Cass, S. (2014). The age of the zettabyte cisco: The future of internet traffic is video dataflow. *IEEE Spectr.* 51, 68. doi:10.1109/mspec.2014.6745894
- Chen, Y., Murherjee, D., Han, J., Grange, A., Xu, Y., Liu, Z., et al. (2018). "An overview of core coding tools in the av1 video codec," in *Picture coding symposium* (San Francisco, CA: PCS), 41–45. doi:10.1109/PCS.2018.8456249
- Everett, H., III (1963). Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Res.* 11, 399–417. doi:10.1287/opre.11.3.399
- Ghanem, B., and Ahuja, N. (2010). "Maximum margin distance learning for dynamic texture recognition," in *European conference on computer vision* (Springer), 223–236.
- Hamza, A. M., Abdelazim, A., and Ait-Boudaoud, D. (2019). Parameter optimization in h 265 rate-distortion by single frame semantic scene analysis. *Electron. Imaging* 262, 262-1–262-6. doi:10.2352/issn.2470-1173.2019.11.ipas-262
- Im, S., and Chan, K. (2015). Multi-lambda search for improved rate-distortion optimization of h.265/hevc. In *2015 10th international conference on information, communications and signal processing (ICICS)*. 1–5. doi:10.1109/ICICS.2015.7459952
- ITU-T RECOMMENDATION, P. (2022). *Subjective video quality assessment methods for multimedia applications*. Geneva, Switzerland: ITU-T, 910.

- Jayant, N. S., and Noll, P. (1984). *Digital coding of waveforms: Principles and applications to speech and video*. Englewood Cliffs, NJ, 115–251.
- Katsavounidis, I., and Guo, L. (2018). “Video codec comparison using the dynamic optimizer framework,” in *Applications of digital image processing XLI* (San Diego, CA: International Society for Optics and Photonics), 10752, 107520Q.
- Kendall, A., Badrinarayanan, V., and Cipolla, R. (2015). *Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding*. *arXiv preprint arXiv:1511.02680*.
- Lin, J. Y., Jin, L., Hu, S., Katsavounidis, I., Li, Z., Aaron, A., et al. (2015). “Experimental design and analysis of jnd test on coded image/video,” in *Applications of digital image processing XXXVIII* (San Diego, CA: International Society for Optics and Photonics), 9599, 95990Z.
- Liu, Z., Mukherjee, D., Lin, W.-T., Wilkins, P., Han, J., and Xu, Y. (2017). Adaptive multi-reference prediction using a symmetric framework. *Electron. Imaging* 2017, 65–72. doi:10.2352/issn.2470-1173.2017.2.vipc-409
- Ma, C., Naser, K., Ricordel, V., Le Callet, P., and Qing, C. (2016). An adaptive Lagrange multiplier determination method for dynamic texture in hevc. In *2016 IEEE international conference on consumer electronics-China (ICCE-China)* (IEEE), 1–4.
- Menon, V. V., Feldmann, C., Amirpour, H., Ghanbari, M., and Timmerer, C. (2022). “Vca: Video complexity analyzer,” in *Proceedings of the 13th ACM multimedia systems conference*, 259–264.
- Mercat, A., Viitanen, M., and Vanne, J. (2020). Uvg dataset: 50/120fps 4k sequences for video codec analysis and development. *Proc. 11th ACM Multimedia Syst. Conf.*, 297–302.
- Mukherjee, D., Bankoski, J., Grange, A., Han, J., Koleszar, J., Wilkins, P., et al. (2013). “The latest open-source video codec vp9-an overview and preliminary results,” in *Picture coding symposium (PCS)* (IEEE), 2013, 390–393.
- Netflix (2015). Netflix open content. Available: <https://opencontent.netflix.com/>.
- Ortega, A., and Ramchandran, K. (1998). Rate-distortion methods for image and video compression. *IEEE Signal Process. Mag.* 15, 23–50. doi:10.1109/79.733495
- Papadopoulos, M. A., Zhang, F., Agrafiotis, D., and Bull, D. (2016). “An adaptive qp offset determination method for hevc,” in *2016 IEEE international conference on image processing (Phoenix, AZ:ICIP)*, 4220–4224. doi:10.1109/ICIP.2016.7533155
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical recipes in C*. Cambridge university press Cambridge.
- Ringis, D. J., Pitié, F., and Kokaram, A. (2021). Near optimal per-clip Lagrangian multiplier prediction in HEVC. In *PCS*. 1–5. doi:10.1109/PCS50896.2021.9477476
- Ringis, D. J., Pitié, F., and Kokaram, A. (2020b). *Per clip Lagrangian multiplier optimisation for (HEVC)*. *Electronic Imaging* 2020.
- Ringis, D. J., Pitié, F., and Kokaram, A. (2020a). “Per-clip adaptive Lagrangian multiplier optimisation with low-resolution proxies,” in *Applications of digital image processing XLIII* (San Diego, CA: International Society for Optics and Photonics), 11510, 115100E.
- Satti, S. M., Obermann, M., Bitto, R., Schmidmer, C., and Keyhl, M. (2019). “Low complexity smart per-scene video encoding,” in *2019 eleventh international conference on quality of multimedia experience (QoMEX)* (IEEE), 1–3.
- Shen, Y., and Kuo, C. (2018). “Two pass rate control for consistent quality based on down-sampling video in HEVC,” in *2018 IEEE international conference on multimedia and expo (ICME)*, 1–6. doi:10.1109/ICME.2018.8486544
- Sullivan, G. J., Ohm, J. R., Han, W. J., and Wiegand, T. (2012). Overview of the high efficiency video coding (hevc) standard. *IEEE Trans. Circuits Syst. Video Technol.* 22, 1649–1668. doi:10.1109/TCSVT.2012.2221191
- Sullivan, G. J., and Wiegand, T. (1998). Rate-distortion optimization for video compression. *IEEE Signal Process. Mag.* 15, 74–90. doi:10.1109/79.733497
- Tourapis, A. M., Singer, D., Su, Y., and Mammou, K. (2017). *BDRate/BD-PSNR Excel extensions*. Joint Video Exploration Team (JVET) of ITU-T and ISO/IEC JTC 1/SC 29/WG 11.
- Vibhoothi, Pitié, F., Katsenou, A., Ringis, D. J., Su, Y., Birkbeck, N., et al. (2022b). “Direct optimisation of λ for HDR content adaptive transcoding in AV1,” in *Applications of digital image processing XLV* (San Diego, CA: SPIE), 1222606. doi:10.1117/12.2632272
- Vibhoothi, Pitié, F., Katsenou, A., Su, Y., Adsumilli, B., and Kokaram, A. (2023). “Comparison of HDR quality metrics in Per-Clip Lagrangian multiplier optimisation with AV1,” in *IEEE international conference on multimedia and expo (ICME) (to be presented)*.
- Vibhoothi, Pitié, F., and Kokaram, A. (2022a). Frame-type sensitive RDO control for content-adaptive encoding. *IEEE Int. Conf. Image Process.*, 1506–1510. doi:10.1109/ICIP46576.2022.9897204
- Wang, Y., Inguva, S., and Adsumilli, B. (2019). “Youtube ugc dataset for video compression research,” in *2019 IEEE 21st international workshop on multimedia signal processing (MMSP)* (IEEE), 1–5. YouTube-UGC dataset. <https://media.withyoutube.com/> (Accessed April 2019).
- Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. In *Thirty-Seventh Asilomar Conf. Signals, Syst. Comput.* 2, 1398–1402.
- Wu, P.-H., Kondratenko, V., Chaudhari, G., and Katsavounidis, I. (2021). “Encoding parameters prediction for convex hull video encoding,” in *2021 picture coding symposium (PCS)* (IEEE), 1–5.
- Wu, P.-H., Kondratenko, V., and Katsavounidis, I. (2020). “Fast encoding parameter selection for convex hull video encoding,” in *Applications of digital image processing XLIII* (San Diego, CA: SPIE), 11510, 181–194.
- Xiph.org (2018). *Xiph.org video test media*. Available at: <https://media.xiph.org/video/derf/>.
- Yang, A., Zeng, H., Chen, J., Zhu, J., and Cai, C. (2017). Perceptual feature guided rate distortion optimization for high efficiency video coding. *Multidimensional Syst. Signal Process.* 28, 1249–1266. doi:10.1007/s11045-016-0395-2
- Zhang, F., and Bull, D. R. (2015). “An adaptive Lagrange multiplier determination method for rate-distortion optimisation in hybrid video codecs,” in *2015 IEEE international conference on image processing (ICIP)* (IEEE), 671–675.
- Zhang, F., and Bull, D. R. (2019). Rate-distortion optimization using adaptive Lagrange multipliers. *IEEE Trans. Circuits Syst. Video Technol.* 1, 3121–3131. doi:10.1109/TCSVT.2018.2873837
- Zhao, X., Lei, Z. R., Norkin, A., Daede, T., and Tourapis, A. (2021). *AOM common test conditions v2.0*. Alliance for Open Media, Codec Working Group Output Document CWG/B075o.