

Per Clip Lagrangian Multiplier Optimisation for HEVC

Daniel J Ringis, François Pitié and Anil Kokaram; Sigmedia Group, Electronic and Electrical Engineering Dept., Trinity College Dublin; Dublin, Ireland

Abstract

The majority of internet traffic is video content. This drives the demand for video compression in order to deliver high quality video at low target bitrates. This paper investigates the impact of adjusting the rate distortion equation on compression performance. A constant of proportionality, k , is used to modify the Lagrange multiplier used in H.265 (HEVC). Direct optimisation methods are deployed to maximise BD-Rate improvement for a particular clip. This leads to up to 21% BD-Rate improvement for an individual clip. Furthermore we use a more realistic corpus of material provided by YouTube. The results show that direct optimisation using BD-rate as the objective function can lead to further gains in bitrate savings that are not available with previous approaches.

Introduction

Video content is predicted to rise to over 80% of internet traffic by the year 2021 [1]. This continues to drive video compression technology. A successful video codec would provide a bitstream which is capable of being decoded at the highest possible quality for a target bitrate, or the lowest possible bitrate for a target quality or distortion. One of the key challenges to be solved in the design of a practical codec is the tradeoff between rate and distortion. A Lagrange multiplier approach was advocated by Sullivan et al [2] and that has been the adopted view since 1998. In this approach the codec makes a number of decisions in order to minimise a cost J as follows.

$$J = D + \lambda R \quad (1)$$

As can be seen J combines both a distortion D (for a frame or macroblock) and a rate R (the number of coded bits for that unit) through the action of the Lagrangian multiplier λ . Different choices for λ result in different R/D tradeoffs. The idea is used across a wide range of codec operations e.g. block type (Skipped/Intra/Inter), motion vectors and bit allocation at the frame and clip levels [3].

The choice of λ was determined experimentally in the early days of codec research [4]. This parameter can be considered as a kind of hyperparameter in a hybrid codec as it affects so many other decisions. This was originally implemented for the H.263 codec and then modified for H.264 and H.265 (HEVC) [5]. The approach was to select a value based on a cohort of examples such that the performance was optimised *on average* across the set. That set was quite small, less than 5 examples. More recently, Katsavounidis et al [6][7] recognised that by choosing codec hyperparameters *per clip* significant gains could be achieved. This is because the statistics of video clips varies greatly over any corpus. While that work considered only the design of a bitrate ladder for DASH streaming, it provided an impetus for *per clip* optimisation throughout the codec system. Covell et al[8] also built on that idea

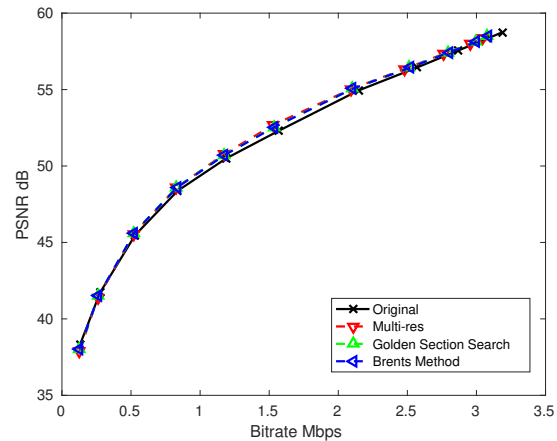


Figure 1. Rate Distortion curves for sequence: Gaming_360P-1e31. The RD curves generated using any one of the proposed algorithms outperforms the curve for X265 using the default λ (black).

by optimising other codec hyperparameters (CRF) in that case for *per clip* transcoding.

This paper explores the idea that a better λ exists for an individual video clip. Unlike previous work, λ was adjusted using direct optimisation techniques using BD-Rate as the objective function. In addition, a more comprehensive and modern corpus of test data was used, provided recently by YouTube [9]. For a given clip it was found that up to a 21% improvement can be had by adjusting the Lagrangian multiplier in the rate distortion equation. An example of the gains reported in this paper are shown in Figure 1. The proposed algorithms all result in RD performance that is well above the default behaviour in the X265 codec (in this case). This result is discussed in more detail later, but it contrasts well with previous work that reports up to 11% improvement. The next section presents more detailed background and a review of recent work.

Lagrangian Multiplier in Compression

There has been a tremendous amount of work done to establish what the Lagrangian multiplier should be in the different iterations of the MPEG codecs (H.263, H.264 and H.265). Being able to find the balance between the quality of decoded images and channel capacity is the fundamental problem which the rate distortion algorithm attempts to solve [10].

As alluded to in the introduction, the seminal work of Sullivan and Wiegand[2] laid the foundation for an experimental approach to choosing an appropriate λ . In that work, they established the relationship between quantisation step size Q (used for DCT coefficient quantisation) and the distortion D in a frame i.e.

$D(Q) = aQ^2$, where a is a constant relating to the content. Minimising J wrt D then yields a relationship $\lambda = D/b$ where b is another content related constant. Combining these two relationships then leads to a relationship between λ and Q as follows

$$\lambda = 0.85 \times Q^2 \quad (2)$$

where the constant was determined experimentally over a small corpus of 3 video clips (CIF resolution). This was explicitly presented for H.263 in 2001 [3].

Updates to these experiments then yielded similar relationships for H.264 and H.265 (HEVC). Because of the introduction of bi-directional (B) frames the constants are all different. In fact three different relationships were established for each of the Intra (I), Predicted (P) and B frames as follows.

$$\lambda_I = 0.57 \times 2^{(Q-12)/3} \quad (3)$$

$$\lambda_P = 0.85 \times 2^{(Q-12)/3} \quad (4)$$

$$\lambda_B = 0.68 \times \max(2, \min(4, (Q-12)/6)) \times 2^{(Q-12)/3} \quad (5)$$

Clearly the size of the corpus used for these experiments was quite small, so there is room for exploration with a larger corpus. In addition, average performance is known to be worse than individual adaptation as discussed previously. Hence the following section outlines different methods which have been used to determine a better Lagrangian multiplier for use in video encoding.

Adaptive Lagrange Multipliers

There have been a few attempts to adjust the rate distortion optimisation within a video codec. In all cases, the authors have discovered that a better λ does exist *per clip*. The techniques are best considered as two classes. In one class (M-Algorithms), λ is adjusted away from the default stated previously, using a constant of proportionality k as follows

$$\lambda_{\text{new}} = k \times \lambda_{\text{orig}} \quad (6)$$

where λ_{orig} is the default Lagrangian multiplier estimated in the video codec, and λ_{new} is the updated Lagrangian. Algorithms in this class, learn a relationship between k and features extracted from the video stream. Zhang and Bull [11] used a single feature D_P/D_B , the ratio between the MSE of P and B frames. This feature gives some notion of temporal complexity. Experiments based on the DynTex database yielded a choice for k as follows.

$$k = a(D_P/D_B + d)^b + c \quad (7)$$

where a, b, c and d were experimentally calculated and different for each codec tested (H.264 and H.265). In their work, k was updated every GOP and they report up to 11% improvement in BD-Rate. This approach of fitting a simple curve was also used by Yang et al, except that they used a different feature, a perceptual content measurement PC as follows

$$k = aPC - b \quad (8)$$

where again a, b were determined experimentally using a corpus of the Derfs dataset. Unlike Zhang and Bull, this was a straight line fit, the loss in complexity of the fit is compensated for by the increase in complexity of the feature.

Ma et al [12] used four features with a SVM to determine k . The SVM was determined using the DynTex dataset. They reported up to 2dB improvement in PSNR and 0.05 improvement in SSIM at equal bitrates. The four video features measured spatial and temporal complexity directly from texture and motion measurements on the pixel level data itself. Hamza et al [13] also take a Machine learning based approach but here using a semantic scene model. Their work used the Derfs dataset and reported up to 6% BD-Rate improvement.

Since λ is linked to Q as explained previously, the second class of algorithms (Q-algorithms) adjusted λ implicitly by adjusting the quantiser parameter Q . Papadopoulos et al [14] applied an offset to Q , in HEVC, based on the ratio of the distortion in the P and B frames. Each QP was updated from the previous Group of Pictures (GOP) using the following equation:

$$QP = a \times \text{MSE}_{\text{ratio}} - b \quad (9)$$

where a, b are constants determined experimentally. This led to an average BD-Rate improvement of 1.07% on the DynTex dataset. With up to 3% BD-Rate improvement achieved for a single sequence. Taking a local, exhaustive approach, Im and Chan [15] proposed encoding a frame multiple times in a video. Each frame was encoded using a Quantiser Parameter from the set ($QP \in (QP, QP \pm 1, QP \pm 2, QP \pm 3, QP \pm 4)$). This led to an increased complexity in coding in HEVC, however for a single sequence achieved a 14.15% BD-Rate improvement.

However, both M/Q-Algorithms have overlooked the possibility that a direct search using BD-Rate as the objective function, could lead to even more improvements. In some way all these previous works have reported data showing k versus $BD - Rate$, but they use various summary models to simplify this relationship across a corpus. By using instead Direct optimisation, we are able to explore just how much more BD-Rate gains there are per-clip. This idea is presented next.

Direct Optimisation

The approach adopted here falls under the M-algorithm class discussed above. Hence an estimate of k is required such that $\lambda_{\text{new}} = k \times \lambda_{\text{orig}}$. In this work however, we directly minimise BD-Rate with respect to k using a range of optimisation strategies. Furthermore, we deploy the optimisation strategies using a modern dataset of User Generated content containing realistic material provided by YouTube [9]. The new objective function is now directly the BD-Rate [16] as a function of $k, B_r(k)$ as follows.

$$B_r(k) = \int_{D_1}^{D_2} (R_1(D) - R_k(D))dD \quad (10)$$

where the integral is evaluated over the quality range $D_1..D_2$. $R_1(D), R_k(D)$ are the RD curves corresponding to the default ($k = 1$) and the evaluated multiplier k respectively. Each RD operating point is generated at a constant bitrate within a range which matches typical streaming media use cases. The overall optimisation process is then as follows.

1. Generate an RD curve using λ_{orig} and $R_t = 256k : 7M$. We use 11 operating points in this range.
2. Define our BD-Rate objective function as specified above in equation 10. We use the same polynomial-log fit for evaluating the integral as recommended in [16].

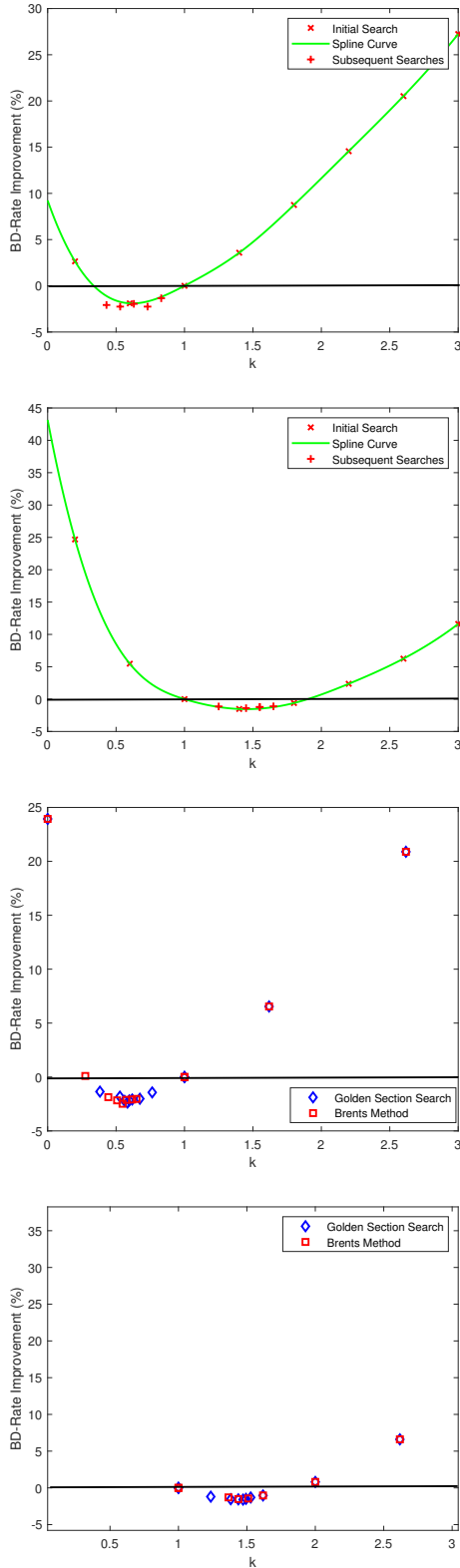


Figure 2. BD Rate(%) vs k using two clips ((NewsClip.720P-7745 and CoverSong.720P-3dca). Top : Multi-Res search, where X's are the initial search results, the solid line is the fit spline interpolation and +s are the subsequent searches. Bottom : Brents Method \diamond and Golden Section \square .

- Starting from $k = 1.0$, minimise $B_r(k)$, BD-Rate, wrt k using any typical optimisation routine.

Note that for every evaluation of $B_r(k)$, eleven (11) encodes are required as well as the subsequent BD-Rate calculation itself.

Optimisation Algorithms

Three simple direct optimisation algorithms were tested. The first used a multi-resolution search (multi-res). This is a simple brute force strategy in which successively refined grid positions in k were searched. The initial grid was $k = 0.2 : 0.4 : 3.0$. The value of k which yielded the minimum $B_r(k)$, k_{opt}^1 was calculated based on a spline curve fit to the sparse points, evaluated at steps of 0.01. Successive refinements searching positions of $k_{\text{opt}}^n \pm \Delta$ were then performed with $\Delta = 0.2, 0.1, 0.05$ in succession. Figure 2 shows this process in action using two clips. X's indicate the initial evaluated positions, and the solid line is the initial curve fit. The subsequent refinement grids are indicated with +. This process therefore always required 15 evaluations of the objective function.

Golden Section Search and Brent's method [17] were used as the other two optimisation schemes. These are smarter direct search methods using local curve fits. As the multi-res method required fifteen (15) evaluations of the objective function, these optimisation routines were terminated after fifteen iterations and a tolerance of 0.02% in the objective function. This allowed a fair comparison.

An example of the search results (k vs BD-Rate) for the methods can be seen in Figure 2 for two particular clips. Implementation details are in the next section, but one key observation here is the very different appearance of each curve. This shows that a summary model (e.g. used by Zhang and Bull) may be unable to exploit all the gains available. Furthermore, we note that the points discovered by the direct search methods deviate from the initial smooth curve fit near the minimum. This shows the complexity of the surface in the region of interest. Finally, note that the initial iterations of Brent's and Golden searches cover a very similar curve as the Multi-res approach. This means that the complexity of the curves is actually limited to the region near the minimum.

Implementation details

For this work VideoLAN's [18] implementation of H.265 was used, x265¹. The codebase was modified to take as an argument, the multiplier k and hence alter the Lagrangian constraint 6 used in RD control throughout the codec.

In order to generate Rate Distortion curves (RD-curve), each video was encoded using the default input parameters for x265 at a constant bitrate. The target bitrate was set within the range of 256k to 7M. Both SSIM and PSNR were used as the distortion metric. The command invocation used for each video encode was as follows:

```
x265 --input SEQ.y4m --bitrate <XXXX>
--tune-<YYYY> --<YYYY> --csv-log-level 2 --csv
DATA.csv --output OUT.mp4
```

Where SEQ, OUT and DATA are the filenames for the raw input file, output encoded video and the log file for an encode re-

¹Version: 3.0+28-gbc05b8a91

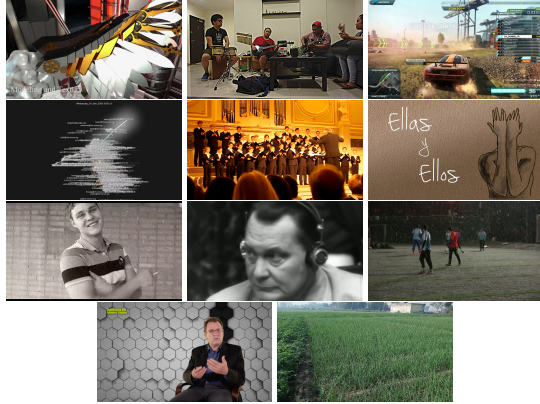


Figure 3. Example frames from each class in the YouTube dataset. Left to right and top to bottom: Animation, CoverSong, Gaming, HowTo, LiveMusic, LyricVideo, MusicVideo, NewsClip, Sports, TelevisionClip, Vlog

spectively. XXXX is the target bitrate and YYYY is the distortion metric used, either PSNR or SSIM.

Dataset

Previous work in this space [11][12][14] has used generally low resolution video with content that does not reflect current use. A sub-sample of videos from the recently released YouTube-UGC Dataset² [9] was therefore selected for use in this paper³. This should provide a better representation of widely used and viewed video content. Each video clip used was 150 frames (5 seconds) in length to match the chunks typically used by Dynamic Adaptive Streaming over HTTP (DASH) systems. The selected clips contained Forty-four (44) at 720p resolution and thirty-three (33) clips at 360p resolution, all at 30 fps. These clips represented eleven (11) classes of video as specified by the YouTube team. A sample of each class of the clips can be seen in Figure 3.

Results

Figure 4 shows the average BD-Rate improvement for each sequence class for each of the three direct optimisation methods. The bar chart shows that in almost all cases, Brent’s and Golden methods are better than multi-res. Per class gains in BD-Rate are within 0.5%-5%. For three classes (Live, Sports and Vlog) multi-res is best. This is probably due to the restrictions on maximum allowed iterations in the competing methods. The best BD-Rate improvement for each class is listed in Table 1. Up to 21% is observed in an animation clip. The corresponding BD-Distortion is less than 0.01dB. This highlights that the gains are all in bitrate without loss in quality, and that there are significant gains to be had *per clip*.

On average golden section and Brent’s search took 12.6 and 9.7 iterations respectively. This was with a 0.02% BD-Rate improvement tolerance used as the stopping criterion. Figure 1 shows a comparison over the RD curves used for a single clip as an example. The gains are larger at lower bitrates as expected. In this example, the BR-Rate gain was 2.5% and Brent’s method took the lowest number of iterations : 8.

²media.withyoutube.com

³The specific location of the chunks in the clips provided is available: <http://www.mee.tcd.ie/~sigmedia/EI2020>

Table 1: Best BD-Rate (%) improvement for each class of sequence using the three direct optimisation methods

Class	Brent	Golden	Multi-Res
Animation	21.667	21.667	13.109
CoverSong	1.839	2.279	2.377
Gaming	1.858	1.762	1.665
HowTo	5.023	5.239	5.037
LiveMusic	1.570	1.810	1.776
LyricVideo	15.879	15.879	10.383
MusicVideo	6.417	6.760	0.516
NewsClip	2.481	2.371	2.251
Sports	8.391	7.583	8.378
TelevisionClip	10.533	10.238	9.486
Vlog	0.351	0.496	0.590

Videos which were encoded without the stopping criterion on the direct optimizer had an additional improvement in BD-Rate. While a further increase in compression performance is welcome, it comes with a heavy computational cost. Figure 5 shows an example of convergence in the direct optimizers in this case without the stringent stopping criterion. As can be seen the BD-Rate gains over 8% after 35 iterations. But the lower plot shows the amount of gains achieved after each iteration as a fraction of the final converged rate. Roughly 80% of the potential maximum BD-Rate improvement can be achieved within the first fifteen(15) iterations, therefore the stopping criterion are not that detrimental to performance especially in the light of computational load.

In practice, we are interested in the rate gains *at a particular* quality. For example a 40dB quality target is a good rule of thumb for a streaming media service. BD-Rate measures instead the average gain across the RD performance curve. We can show our results in this light. Figure 6 is a cumulative histogram of the percent bitrate improvement at 40 dB for the 77 clips encoded in this paper. For 80% of the clips we can have up to 5% improvement at this quality. This is a useful metric and further validates the proposed methods.

Table 2 compares the results from this work with the results from Zhang and Bull [11] applied to the dataset in this paper. The comparison is not strictly fair because we do not adapt the k every GOP and instead use a single GOP for all the clips. However the results further reiterate the difference in the shapes of the curves *per clip*. As can be seen our proposed algorithm estimates values of k that are quite different, and subsequently the BD-Rate improvement is much better than that using k as equation 7.

Conclusion

This paper has presented a direct optimisation technique for Lagrange multiplier estimation in a codec. The approach in some sense represents an upper bound for what can be achieved by adjusting the Lagrangian multiplier. We report gains up to 21% on a particular clip and 0.5 – 5% per class without affecting the quality. At a target quality of 40 dB, 80% of the clips had gains up to 5%. Furthermore, the work here has used a much more realistic database of content provided by YouTube and for a larger number of clips. The results also show that the relationship between k and BD-Rate varies substantially between clips, and that the region

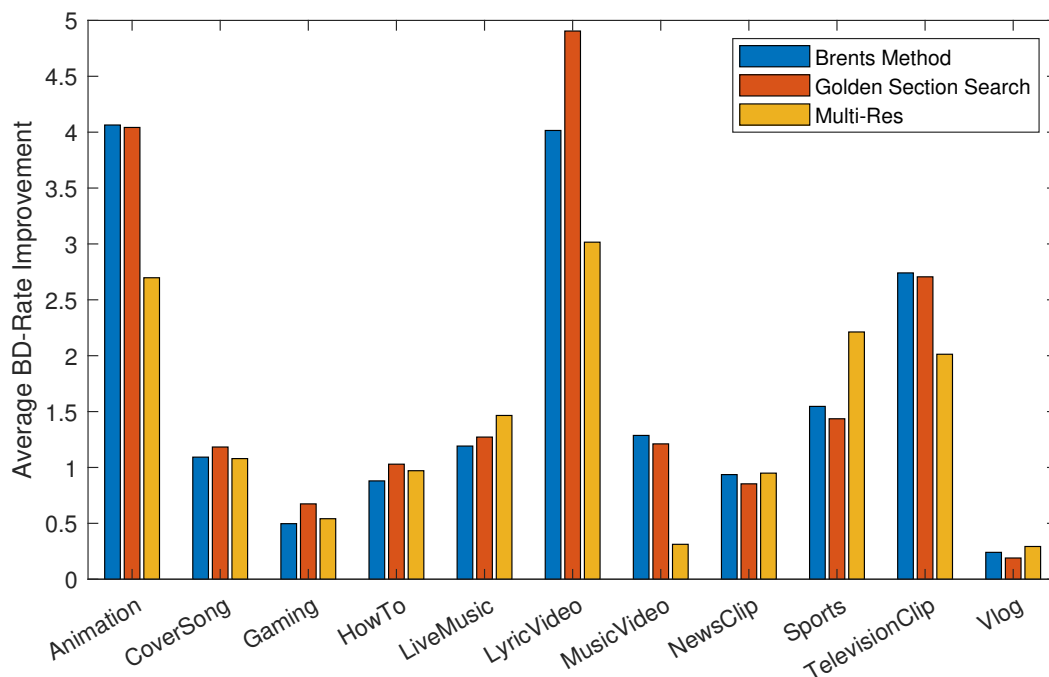


Figure 4. Average BD-Rate improvement (%) for each class of video. Both Brent's Method and Golden Section Search require less iterations than Multi-Res search and often provide better BD-Rate improvement.

Table 2: Comparison of the BD-Rate improvement using an estimated k from [11] vs k determined by our direct search (Brent's method). As can be seen our estimated values are quite different and lead to much improved performance. However our adaptation scheme is different from that in Zhang and Bull. Negative BD-R implies worse performance than the unmodified codec

Clip	k [11]	BD-R	k	BD-R
Animation4268	3.173	-5.81	0.777	0.591
CoverSong7360	3.444	-4.23	1.071	0.178
Gaming6403	3.132	-6.18	1.438	0.683
HowTo21c6	2.938	-3.15	0.438	5.239
LiveMusic6452	3.632	-0.85	1.145	0.234
LyricVideo739a	2.491	0.89	1.699	1.412
MusicVideo3285	4.298	-10.21	1.120	0.068
NewsClip7745	2.571	-2.67	0.547	2.482
Sports0b9e	3.108	1.89	1.784	8.391
TelevisionClip02b8	2.112	-3.12	0.587	2.887
Vlog11c5	3.579	-2.12	1.096	0.166

near the minimum is not a simple surface. These observations point to further work in *per clip* optimisation using more sophisticated machine learning techniques that exploit deeper features in the content.

References

- [1] S. Cass, "The age of the zettabyte cisco: the future of internet traffic is video [dataflow]," *IEEE Spectrum*, vol. 51, no. 3, pp. 68–68, 2014.
- [2] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE signal processing magazine*, vol. 15, no. 6, pp. 74–90, 1998.
- [3] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 3, pp. 542–545, IEEE, 2001.
- [4] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal processing magazine*, vol. 15, no. 6, pp. 23–50, 1998.
- [5] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [6] M. M. J. D. C. A. Aaron, Z. Li and D. Ronca, "Netflix Technology Blog - Per-Title Encode Optimization." <https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2>.
- [7] I. Katsavounidis and L. Guo, "Video codec comparison using the dynamic optimizer framework," in *Applications of Digital Image Processing XLI*, vol. 10752, p. 107520Q, International Society for Optics and Photonics, 2018.
- [8] M. Covell, M. Arjovsky, Y.-c. Lin, and A. Kokaram, "Optimizing transcoder quality targets using a neural network with an embedded bitrate model," *Electronic Imaging*, vol. 2016, no. 2, pp. 1–7, 2016.
- [9] Y. Wang, S. Inguva, and B. Adsumilli, "Youtube ugc dataset for video compression research," *arXiv preprint arXiv:1904.06457*, 2019.
- [10] 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-*

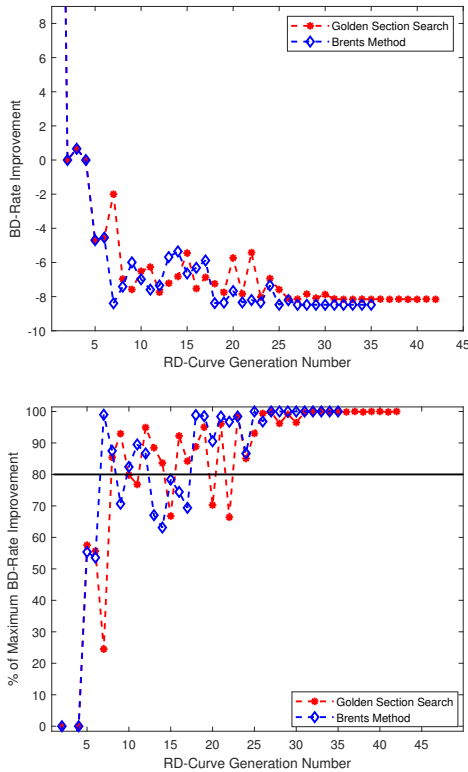


Figure 5. These two plots illustrate the convergence of the optimisation methods used for a particular clip (Sports_720P-0b9e). Top : BD-Rate. Bottom : Normalised BD-Rate improvement. The normalised plot shows the BD-Rate improvement as a % of the maximum BD-Rate possible for this clip. Hence it shows that 80% of the possible BD-Rate improvement is attained within the first 15 iterations.

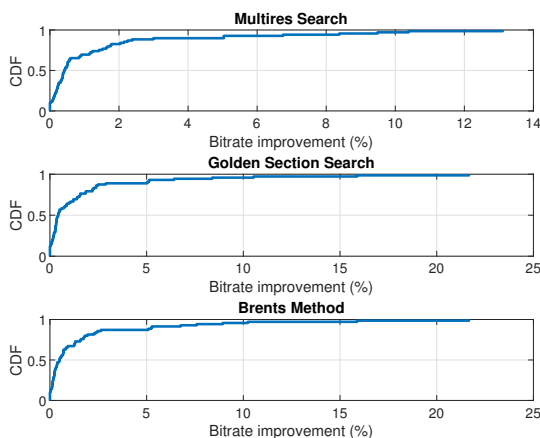


Figure 6. CDF vs bitrate improvement (%) at 40dB. Top to bottom: Multires, Golden Section Search, Brents Method. At a target quality of 40 dB, 80% of the clips had gains up to 5% improvement for Golden Section and Brents Method

actions on Circuits and Systems for Video Technology, vol. 6, no. 2, pp. 182–190, 1996.

[11] F. Zhang and D. R. Bull, “Rate-distortion optimization using adaptive lagrange multipliers,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2018.

[12] C. Ma, K. Naser, V. Ricordel, P. Le Callet, and C. Qing, “An adaptive lagrange multiplier determination method for dynamic texture in hevc,” in *2016 IEEE International Conference on Consumer Electronics-China (ICCE-China)*, pp. 1–4, IEEE, 2016.

[13] A. M. Hamza, A. Abdelazim, and D. Ait-Boudaoud, “Parameter optimization in h. 265 rate-distortion by single frame semantic scene analysis,” *Electronic Imaging*, vol. 2019, no. 11, pp. 262–1, 2019.

[14] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. Bull, “An adaptive qp offset determination method for hevc,” in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 4220–4224, Sep. 2016.

[15] S. Im and K. Chan, “Multi-lambda search for improved rate-distortion optimization of h.265/hevc,” in *2015 10th International Conference on Information, Communications and Signal Processing (ICICSP)*, pp. 1–5, Dec 2015.

[16] G. Bjontegaard, “Calculation of average psnr differences between rd curves; vceg-m33,” tech. rep., 2001.

[17] B. P. Flannery, W. H. Press, S. A. Teukolsky, and W. Vetterling, “Numerical recipes in c,” *Press Syndicate of the University of Cambridge, New York*, vol. 24, p. 78, 1992.

[18] V. Organization, “x265, the free H.265/HEVC encoder.” <https://www.videolan.org/developers/x265.html>.

Author Biography

Daniel J. Ringis is a PhD researcher at Trinity College Dublin. He is from Trinidad and Tobago, and previously studied Electrical and Computer Engineering at the University of the West Indies, St. Augustine. His current research is focused on video compression.

François Pitié is an Ussher Assistant Professor in Media Signal Processing in the School of Engineering and part of the ADAPT Research Centre. His research is concerned with media signal processing applications related to film postproduction and film restoration. The aim of his research is to develop tools that can enhance the cinematic experience of immersion.

Anil Kokaram received the Ph.D. degree from Cambridge University, U.K., in 1993. Then, he was a Research Fellow with the Cambridge University Engineering Department and Churchill College Cambridge. In 1998, he established the Sigmedia Group (www.sigmedia.tv) at Trinity College Dublin. In 2011 a company he founded, Green Parrot Pictures, operating in the consumer image processing space, was acquired by Google. He is currently working as a Tech. Lead in the Chrome Media Group at Google, Mountain View CA. He continues to hold his Professorship with the Department of Electronic Engineering, Trinity College Dublin and is a Fellow of the College. He is the author of over 100 refereed papers in various conferences and journals over the years. His research interests include the general area of digital video processing and algorithms for stereoscopic quality improvement and video processing in the life sciences. He is best known for his work in automated motion picture restoration.

JOIN US AT THE NEXT EI!

IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

www.electronicimaging.org

