



Interframe-Dependent Rate-QP-Distortion Model for Video Coding and Transmission

Mourad Aklouf, Marc Leny, Michel Kieffer, Frédéric Dufaux

► To cite this version:

Mourad Aklouf, Marc Leny, Michel Kieffer, Frédéric Dufaux. Interframe-Dependent Rate-QP-Distortion Model for Video Coding and Transmission. IEEE International Conference on Image Processing (ICIP'2021), Sep 2021, Anchorage, United States. pp.2019-2023, 10.1109/ICIP42928.2021.9506326 . hal-03231445

HAL Id: hal-03231445

<https://hal.science/hal-03231445>

Submitted on 14 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INTERFRAME-DEPENDENT RATE-QP-DISTORTION MODEL FOR VIDEO CODING AND TRANSMISSION

Mourad AKLOUF^{*†}

Marc LENY^{*}

Michel KIEFFER[†]

Frédéric DUFAUX[†]

^{*}EKTACOM, F-91940 Les Ulis

[†]Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, F-91192 Gif-sur-Yvette

[†]{mourad.aklouf, michel.kieffer, frederic.dufaux}@l2s.centralesupelec.fr, *mleny@ektacom.com

ABSTRACT

In this paper, we propose a new inter-dependent Rate-QP-Distortion model. This model predicts the size of the picture after compression based on the distortion (D) of the reference frame and the current Quantization parameter (QP). This model is particularly useful when adjusting the QP of the picture according to the allocated bitrate budget. Simulation results demonstrate that the proposed model outperforms other models in the literature. In the video sequence Tango, up to 90% of all prediction errors are inferior to 8.6% when using constant QP encoding, and 90% of all prediction errors are inferior to 12% when using variable QP encoding. One application of this model is in low latency video streaming, where each frame of the video sequence needs to be coded with a specific target bitrate, due to variations of the instantaneous transmission rate.

Index Terms— HEVC, x265, inter-frame dependency, Rate-QP model, video coding.

1. INTRODUCTION

In video coding, the control of the encoding rate is of paramount importance to meet strict rate constraints imposed by band-limited transmission channels. This is useful when encoded videos have to be transmitted using, for instance, 5G evolved Multimedia Broadcast & Multicast Services (eMBMS) [1], although buffers at the transmitter and at the receiver may mitigate encoding and transmission rate jitter. When ultra-low latency live streaming is required, such as in remote driving [2] or remote surgery applications [3], rate control becomes of paramount importance. In such applications, a significant part of the delay is introduced by the video encoder and decoder, whereas only very small buffers are allowed to limit the delay to a minimum. Consequently, any mismatch between the encoding rate and available bandwidth on the transmission channel may lead to a unacceptable increase of the delay.

Video encoding rate control usually relies on a model of the dependency between the size of the bitstream resulting from the encoding of a single or a group of frames, and the video encoding parameters. The quantization parameters QP at the frame level or at the smallest coding unit level are usually considered to adjust the encoding rate.

Parametric models between the rate and QP have been proposed in the literature, some of them are reviewed in Section 2. Usually, a trade-off has to be found between the accuracy and the complexity of the model, *i.e.*, the number of parameters involved in the model structure. The model complexity impacts directly that of the parameter estimation process which has to be performed while encoding the video sequence. These models also differ by the type of information

they use to infer the rate (and quality) of the resulting encoded frame. Some information are relatively high-level and do not need a precise knowledge of the behavior of the encoder, *e.g.*, the distortion of the reference frame. Some others are evaluated within the encoder during the encoding process, *e.g.*, the Lagrangian multiplier λ and the percentage ρ of zero-valued transformed coefficients.

Preliminary experiments allow us to observe that the relation between the rate R_n and the quantization parameter QP_n of the n -th frame depends significantly on the distortion of the corresponding reference frames. In this paper, assuming a low-delay encoding profile, where the reference frame used to encode the n -th frame is assumed to be the previous one, we propose a new model of the relation between R_n and QP_n depending on the Mean Square Error (MSE) distortion D_{n-1} for the reference frame $n - 1$. This model, denoted as R-(QP,D), is especially suitable for video coding and low-latency rate adjustment.

After reviewing some related models to predict the encoding rate of video frames in Section 2, Section 3 introduces the proposed frame-level R-(QP,D) model. We especially give some insights on the way the structure of the model has been obtained. Then, Section 4 compares the performance of the proposed approach with state-of-the-art models in terms of accuracy of the encoding rate prediction, especially when the encoding parameters change significantly with time, as required by ultra low-latency streaming applications. Finally, we draw some conclusions in Section 5.

2. RELATED WORKS

The relation between the picture bitrate and the QP is widely discussed in literature. Models can be classified into two categories. The first category assumes that the coding units are independent of each others, whereas the second category considers dependencies among the coding units.

In [4], a model between the rate R of a frame and the quantization step size Q is proposed. This model has been refined in [5] to predict the encoding rate of the H.265/HEVC frames and CTUs, accounting for the Mean Absolute Difference (MAD) between the original $p_k(i, j)$ and the reconstructed k -th coding unit $\hat{p}_k(i, j)$ of size $M \times N$, $MAD_k = \sum_{i=1}^M \sum_{j=1}^N |\hat{p}_k(i, j) - p_k(i, j)|$

$$R_k = M \cdot N \cdot MAD_k \cdot \left(\frac{p_1}{Q_k^2} + \frac{p_2}{Q_k} \right), \quad (1)$$

where $Q_k = 2^{\frac{QP_k - 4}{6}}$. The MAD of CTU k is predicted using that of the CTU $k - 1$: $MAD_k = p_3 MAD_{k-1} + p_4$. This model involves thus 4 parameters, p_1 , p_2 , p_3 and p_4 .

In the same spirit, [6] and [7] present a model involving the Sum of Absolute Difference (SAD) to describe the rate of the k -th frame

$$R_k = p_1 \frac{SAD_k}{Q_k} + p_2, \quad (2)$$

where $SAD_k = M \cdot N \cdot MAD_k$. Additionally, SAD_k is inferred as $SAD_k = SAD_{k,k-1}^{org} + p_3 \sqrt{D_{k-1}} + p_4$, where $SAD_{k,k-1}^{org}$ denotes the SAD between the original frame k and the original reference frame $k-1$. D_{k-1} is the MSE of the reference frame after encoding, $D_{k-1} = \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N (\hat{p}_{k-1}(i, j) - p_{k-1}(i, j))^2$. This model also involves 4 parameters.

The models introduced in [5] and [7] only account for the quality of the reference CTU $k-1$ via its SAD or its MAD. This brings us to the second category of R-Q¹ models [8], where the temporal dependency between the H.264/AVC macroblocks is more explicitly taken into account to get the model

$$R_k = p_1 \cdot M \cdot N \frac{\sigma_k^2}{Q_k^2} \quad (3)$$

where σ_k is the standard deviation of the motion-compensated residue of the k -th macroblock. This model has a single parameter. All above models have been designed to adjust the QP on the block level, but they may be extended at the whole frame level.

Lin and Ortega [9] propose a piecewise cubic model of the relation between R_k of frame k , QP_k and the QP of the reference frame. However, Zhang et al. [10] experimentally show that the piecewise cubic model is inaccurate in fine-granular scalability (FGS) coders.

Two other types of models have also been proposed. The R- ρ model [11] predicts the rate R using the percentage ρ of zero-valued transformed coefficients. Finally, the R- λ [12] predicts the rate R using the Lagrangian multiplier λ . These two approaches model the rate using low-level encoding parameters. Thus, they do not provide a straightforward control of the encoder behavior as the R-Q models. This can represent a weakness in applications such as low latency video streaming. Consequently, only R-Q models are compared against our proposed model in this paper.

3. INTER-DEPENDENT R-(QP,D) MODEL

This section introduces the proposed R-(QP,D) model for rate control at the frame level. The proposed model has been derived based on experiments with H.265/HEVC [13] (using the x265 encoder [14]), Nevertheless, the methodology is generic and can be extended to VVC [15] or AV1 [16].

In order to motivate our model, we show experiments considering two typical frames, with indexes $n = 79$ and $n = 131$ of the video sequence *ParkScene*² encoded with x265 [14]. Figure 1-a shows the rate R_n of frame n as function of the distortion of the reference frame D_{n-1} , for different values of QP_n . These results have been obtained by encoding frame n with six different $QP_n \in \{QP^{(1)} = 20, QP^{(2)} = 24, \dots, QP^{(6)} = 40\}$.

All previous frames, including frame $n-1$ have been encoded with seven $QP_{n-1} = QP_n + \Delta QP$, where $\Delta QP \in \{\Delta QP^{(1)} = -7, \dots, \Delta QP^{(7)} = 5\}$. The resulting rates are denoted as $R_{n,i,j}$, where $QP_n = QP^{(i)}, i = 1, \dots, 6$ and $\Delta QP = \Delta QP^{(j)}, j = 1, \dots, 7$. Each curve for a given value of QP_n consists of two almost linear parts as a function of $\log(D_{n-1})$. Conversely, for small values of D_{n-1} , R_n increases slowly with D_{n-1} . For larger values of D_{n-1} , the increase is steeper. We observe that the dependance

¹This notation covers both R-Q and R-QP models, as there is a direct relationship between the quantization step size Q and QP .

²Experiments performed on other sequences show similar results but are not shown due to space limit.

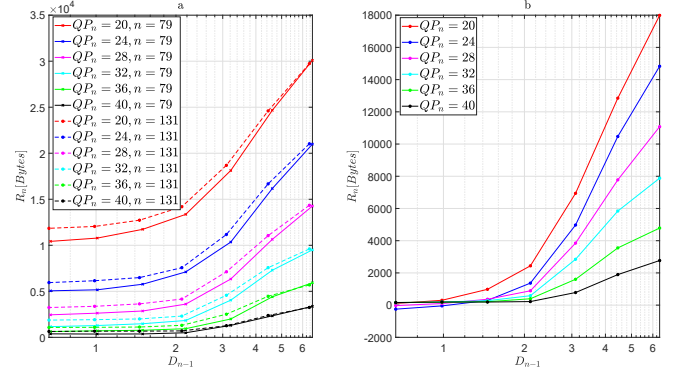


Fig. 1. (a) The bitrate R_n for the frames $n = 79$ and $n = 131$ of *ParkScene* as a function of the distortion D_{n-1} of the reference frame for different values of QP_n ; (b) R_n^0 for frame $n = 79$ of *ParkScene* as a function of D_{n-1} for different values of QP_n

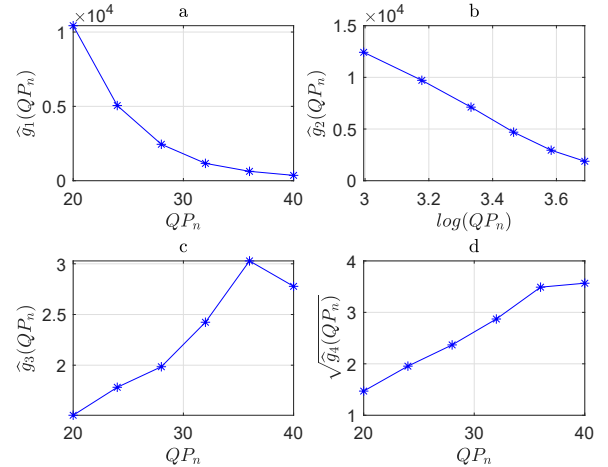


Fig. 2. \hat{g}_1 , \hat{g}_2 , \hat{g}_3 , and \hat{g}_4 as a function of $QP^{(i)}$ or $\log(QP^{(i)})$ for frame 79 of *ParkScene*

of $R_{n,i,j}$ can be described by a family of sigmoids depending on $\log(D_{n-1})$ and QP_n .

We propose the following R-(QP,D) model

$$R_n(QP_n, D_{n-1}) = g_1(QP_n) + g_2(QP_n) (\tanh(g_3(QP_n) \log(D_{n-1}) - g_4(QP_n)) + 1), \quad (4)$$

where $g_1(QP_n)$ describes the bitrate R_n for small values of D_{n-1} , and

$$R_n^0(QP_n, D_{n-1}) = R_n(QP_n, D_{n-1}) - g_1(QP_n) = g_2(QP_n) (\tanh(g_3(QP_n) \log(D_{n-1}) - g_4(QP_n)) + 1), \quad (5)$$

describes the bitrate R_n for large values of D_{n-1} .

Figure 2-a represents R_n as function of QP_n for the smallest values of D_{n-1} obtained for frame $n = 79$. In this regime, R_n decreases exponentially with QP_n according to

$$g_1(QP_n) = p_1 \exp(-p_2 QP_n). \quad (6)$$

Figure 1-b represents $R_{n,i,j}^0 = R_{n,i,j} - g_1(QP^{(i)})$ as a function of D_{n-1} for different values of $QP_n = QP^{(i)}$. For each value of

$QP_n = QP^{(i)}$, $i = 1, \dots, 6$, a least-squares estimation of g_2 , g_3 , and g_4 is performed using $R_{n,i,j}^0$, $j = 1, \dots, 7$ to get $\hat{g}_2(QP^{(i)})$, $\hat{g}_3(QP^{(i)})$, and $\hat{g}_4(QP^{(i)})$. Figure 2-b shows that \hat{g}_2 as a function of $\log(QP^{(i)})$ is adequately described by an affine model with two parameters p_3 and p_4

$$g_2(QP_n) = p_3(-p_4 \log(QP_n) + 1). \quad (7)$$

Figure 2-c illustrates \hat{g}_3 as function of $QP^{(i)}$, which is adequately represented by the linear model depending on p_5

$$g_3(QP_n) = p_5 QP_n. \quad (8)$$

Finally, Figure 2-d illustrates the relation between the square root of \hat{g}_4 and $QP^{(i)}$, which justifies the following quadratic model for the dependency of g_4 in QP_n , depending on the parameters p_6 and p_7

$$g_4(QP_n) = (p_6 QP_n - p_7)^2. \quad (9)$$

Consolidating the previous results, the proposed model (4) involves a vector of 7 parameters $\mathbf{p} = (p_1, \dots, p_7)$, which value is frame-dependent and needs to be determined to accurately predict R_n as a function of QP_n and D_{n-1} .

4. EVALUATION OF THE PROPOSED MODEL

The performance of the proposed model to predict R_n as a function of QP_n is compared to the reference models (1) [5], (2) [7], and (3) [8], used at a frame level.

4.1. Experimental setup

Three JCT-VC test sequences, namely *Tango*, *Racehorses*, and *ParkScene* [17], are selected for the experiments, as well as a recording from the inside of racing car (*Magnycours*). The encoding is performed with the x265 software [14], configured in low delay mode and with an intra-refresh cycle of one second.

The parameters for the three reference models and the proposed models are only estimated for every four frames of the video sequences. The parameters are then assumed to remain constant for the next three frames. For that purpose, 42 coding trials are done with $QP_n \in \{QP^{(1)}, \dots, QP^{(6)}\}$ and $QP_{n-1} = QP_n + \Delta QP$ with $\Delta QP \in \{\Delta QP^{(1)}, \dots, \Delta QP^{(7)}\}$ to get $R_{n,i,j}$, $i = 1, \dots, 6$, and $j = 1, \dots, 7$. A weighted least-square estimation of the model parameters is then performed considering the following cost function

$$C_n(\mathbf{p}) = \sum_{i=1}^6 \sum_{j=1}^7 \frac{1}{R_{n,i,j}} \left(R_{n,i,j} - R_n(QP^{(i)}, D_{n-1}^{(j)}) \right)^2, \quad (10)$$

where $R_n(QP^{(i)}, D_{n-1}^{(j)})$ is given by the proposed model in (4), or by the models (1), (2) or (3), and $D_{n-1}^{(j)}$ is the distortion for frame $n-1$.

To compare the performance of the four models, in a first set of experiments all frames of the video sequences are encoded at constant $QP \in \{QP^{(1)}, \dots, QP^{(6)}\}$. In a second set of experiments, QP may vary from frame to frame as the realization of a first-order Markov process such that with a probability $P = 0.6$, $QP_n = QP_{n-1}$, and with a probability $1 - P$, QP_n is uniformly distributed in the set $\mathcal{Q}_n = \{QP_{n-1} - 5, \dots, QP_{n-1} + 5\} \cap \{20, \dots, 40\}$.

The ability to predict the actual encoding rate is evaluated using the relative rate error

$$E_n = 100 (R_{\text{pred}} - R_{\text{actual}}) / R_{\text{actual}}, \quad (11)$$

where R_{actual} is the actual size of the encoded frame n and R_{pred} is the predicted one obtained from the rate model.

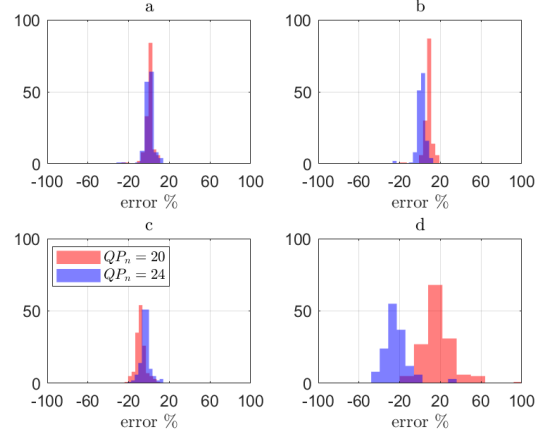


Fig. 3. Histogram of prediction errors for *Tango* at high bitrates, (a) proposed model (4), (b) (1) from [5], (c) (2) from [7] and (d) (3) from [8].

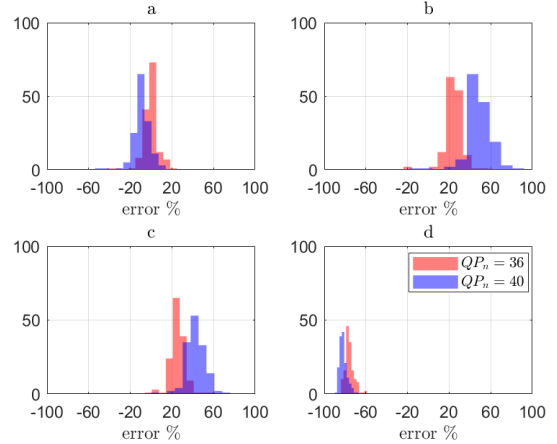


Fig. 4. Histogram of prediction errors for *Tango* at low bitrates, (a) proposed model (4), (b) (1) from [5], (c) (2) from [7] and (d) (3) from [8].

4.2. Results

We first compare the performance of the four models in a set of experiments when encoding with constant QP .

Figure 3 shows the histogram of the prediction errors obtained with the proposed model in (4), and by the models in (1), (2) and (3), used on the *Tango* sequence coded at high bitrates, i.e., $QP_n = 20, 24$. We notice that our proposed model provides the best performance in this case. The errors of model (4) are mainly between -13.6% to 14%, with a peak at 3.4%. The prediction errors with the models (1) and (2) are between -11% and 19%, with a peak near 11% and -6.5% respectively. Model (3) has the worst performance, with prediction errors spreading between -50% and 64%.

Figure 4 shows the error histograms of the four models on the *Tango* sequence coded at low bitrates, i.e., $QP_n = 36, 40$. The performance of our proposed model slightly decreases but significantly outperforms the three other ones. Prediction errors are between -28.6% and 18.3%, with a peak around 4.9%. Both models in (1) and (2) lead to errors between 0% and 81% with a peak near 48% and 27.5% respectively. The model in (3) largely underestimates the rate, with prediction errors distributed between -90% and -58%.

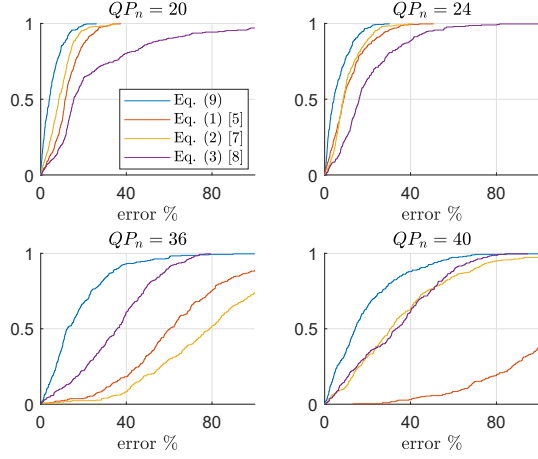


Fig. 5. CDF of prediction errors for *Magnycours* sequence.

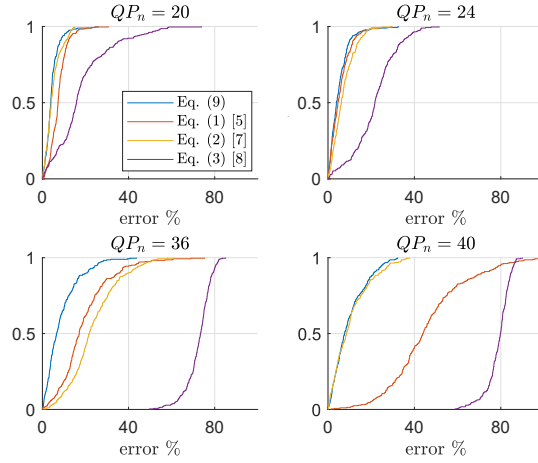


Fig. 6. CDF of prediction errors for *RaceHorses* sequence.

Figure 5 shows the error cumulative distribution functions (CDF) when using the four $QP_n = 20, 24, 36$ and 40 for the *Magnycours* sequence. The proposed model achieves the lowest prediction error compared to the other models. For instance, with $QP_n = 20$, 90% of the prediction errors are less than 12.2% with the proposed model, compared to 16.7%, 22.6%, and 51.8% for the models in (1), (2), and (3), respectively.

Figure 6 shows the error CDF for *RaceHorses* sequence. Here, we see close performance of models (4), (1), and (2) with $QP_n = 20$ and 24 , and close performance between model (4) and (1) with $QP_n = 40$. The proposed model shows a significant advantage in the other cases.

Figures 5 and 6 both show that the gains with our model tend to be more significant at low bitrates (*i.e.*, high QPs). This can be explained by the fact that we incorporate the distortion of the reference frame in our model, which is especially important at low bitrates.

Figure 7 illustrates the average error CDF when coding the four sequences with a constant QP . Our proposed model achieves the best performance for all test sequences.

Next, we present results of the second set of experiments using time-varying QPs based on a first-order Markov process. Figure 8 shows the average error CDF with time-varying QPs for the four test

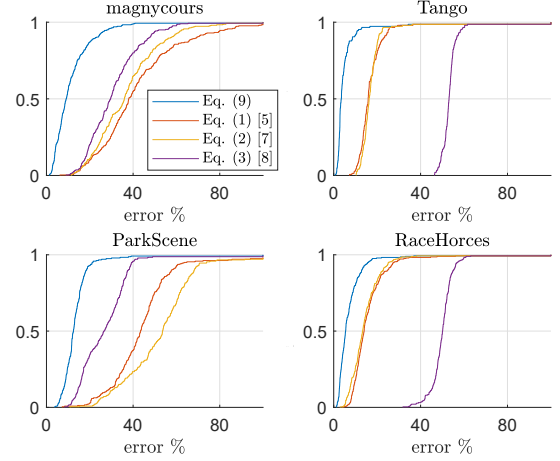


Fig. 7. Average error CDF with constant QP for each sequence.

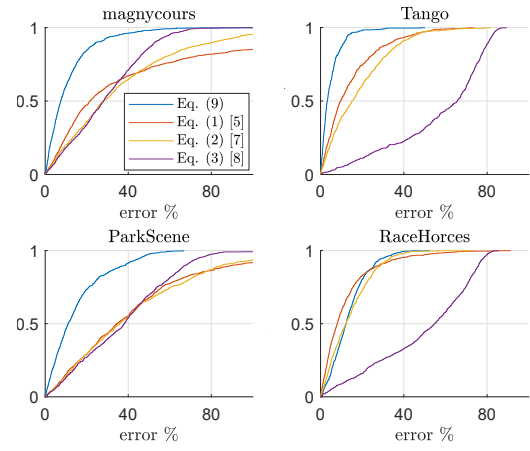


Fig. 8. Error CDF with first-order Markov process variations of QP for each sequences, (a) proposed model (4), (b) (1) from [5], (c) (2) from [7], and (d) (3) from [8]

sequences. The proposed model outperforms the other ones for all sequences. For *Magnycours*, *ParkScene*, and *Tango*, the improvements are significant, whereas for *RaceHorses*, the models in (4), (1), and (2) reach close performances.

5. CONCLUSION

This paper presents a new inter-dependent Rate-QP-Distortion model to predict the size of the n -th encoded frames based on the distortion D_{n-1} of the reference frame and the current quantification parameters QP_n . The performance of the proposed model is evaluated in two coding scenarios: encoding at constant QP and encoding with frame-dependent QP. In both cases, the proposed model outperforms other models in the literature. The gains are especially significant at low bitrates (high QPs), showing the benefits of accounting for the distortion of the reference frame in our model. Finally, the accuracy of the prediction relies on the quality of the parameters fitting and on the characteristics of the video sequence. In our future work, we will focus on the integration of the proposed model in a rate control algorithm for low-delay video streaming.

6. REFERENCES

- [1] Wei Guo, Manuel Fuentes, Louis Christodoulou, and Belkacem Mouhouche, "Roads to multimedia broadcast multicast services in 5G new radio," in *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2018, pp. 1–5.
- [2] Lei Kang, Wei Zhao, Bozhao Qi, and Suman Banerjee, "Augmenting self-driving with remote control: Challenges and directions," in *Proc. of the 19th International Workshop on Mobile Computing Systems & Applications*, 2018, pp. 19–24.
- [3] Sourabh Khire, Scott Robertson, Nikil Jayant, Elena A Wood, Max E Stachura, and Tamer Goksel, "Region-of-interest video coding for enabling surgical telementoring in low-bandwidth scenarios," in *Proc. IEEE Military Communications Conference*, 2012, pp. 1–6.
- [4] Wei Ding and Bede Liu, "Rate control of mpeg video coding and recording by rate-quantization modeling," *IEEE transactions on circuits and systems for video technology*, vol. 6, no. 1, pp. 12–20, 1996.
- [5] Hyomin Choi, Jonghun Yoo, Junghak Nam, Donggyu Sim, and Ivan V Bajić, "Pixel-wise unified rate-quantization model for multi-level rate control," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1112–1123, 2013.
- [6] Siwei Ma, Wen Gao, and Yan Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE transactions on circuits and systems for video technology*, vol. 15, no. 12, pp. 1533–1544, 2005.
- [7] Yuan Li, Huizhu Jia, Pan Ma, Chuang Zhu, Xiaodong Xie, and Wen Gao, "Inter-dependent rate-distortion modeling for video coding and its application to rate control," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2014, pp. 1–6.
- [8] Xiaokang Yang, Yongmin Tan, and Nam Ling, "Rate control for H.264 with two-step quantization parameter determination but single-pass encoding," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 063409, 2006.
- [9] Liang-Jin Lin and Antonio Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 446–459, 1998.
- [10] Xi Min Zhang, Anthony Vetro, Yun Q Shi, and Huifang Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 2, pp. 121–130, 2003.
- [11] Meng Liu, Yi Guo, Houqiang Li, and Chang Wen Chen, "Low-complexity rate control based on ρ -domain model for scalable video coding," in *Proc. IEEE International Conference on Image Processing*, 2010, pp. 1277–1280.
- [12] Minhao Tang, Jiangtao Wen, and Yuxing Han, "A generalized rate-distortion- λ model based HEVC rate control algorithm," *arXiv preprint arXiv:1911.00639*, 2019.
- [13] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas 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.
- [14] MulticoreWare, "x265 software documentation," <https://x265.readthedocs.io/en/master/>, 2020.
- [15] Joint Video Exploration Team, "Versatile video coding (VVC)," <https://jvet.hhi.fraunhofer.de/>, 2019.
- [16] AOMedia Codec, "AV1," <https://github.com/AOMediaCodec/av1-spec>, 2019.
- [17] Teruhiko Suzuki, "JVET-d1002 : Work plan for assessment of test material," *JVET 4th Meeting, Chengdu, CN.*, 2016.