Two-Pass Rate Control for Improved Quality of Experience in UHDTV Delivery

Zupancic, I; Naccari, M; Mrak, M; Izquierdo, E

Two-Pass Rate Control for Improved Quality of Experience in UHDTV Delivery

Ivan Zupancic, *Student Member, IEEE*, Matteo Naccari, *Member, IEEE*, Marta Mrak, *Senior Member, IEEE*, and Ebroul Izquierdo, *Senior Member, IEEE*

***Abstract*—Rate control plays an important role in any video coding application and it was extensively studied in the context of previous video coding standards. However, the current state-of-the-art high efficiency video coding (HEVC) standard introduces many flexible tools making previous rate-distortion models used in rate control insufficiently accurate. Recently, a few rate control methods have been developed for HEVC that introduce many useful features, such as a robust correspondence between the rate and Lagrange multiplier λ. Nonetheless, previous rate control algorithms for HEVC do not address typical content in television applications that consists of frequent scene changes. Furthermore, the new ultra high definition television (UHDTV) format, which is expected to become widespread in the future, demands for even higher compression efficiency. To overcome these issues, a two-pass rate control method is proposed in this paper, targeting the encoding of UHDTV content. In the first pass, a fast encoder with limited set of coding tools is used during pre-encoding step to obtain the data used for rate allocation and model parameter initialization, which will then be used during the second pass. To avoid multiple encoding steps when deriving this information, a variable quantization parameter framework is proposed. Experimental results show that the proposed rate control method outperforms the well-known HEVC rate control method. When compared with variable bit-rate encoding mode, the proposed two-pass rate control method achieves on average 2.9% BD-rate losses. That is significantly better than the state-of-the-art HEVC rate control method, which achieves an average 8.8% BD-rate loss. The proposed method also provides a more consistent quality fluctuation with time, measured with standard deviation of frame PSNR values, required for high Quality of Experience.**

***Index Terms*—HEVC, quality of experience, rate control, UHD video, video streaming.**

I. Introduction

ULTRA high definition television (UHDTV) is the new format which is expected to deliver a greater impact, more presence and immersion than the current high definition television (HDTV). UHDTV is not just about more pixels but it has the potential to deliver wider color gamut, high dynamic range and high frame rate; in other words, it will ultimately provide users with *better* pixels. The parameters for UHDTV are specified in the ITU Recommendation BT.2020 [1] where two spatial resolutions are standardized: 3840×2160 luma samples/frame and 7680×4320 luma samples/frame, both of which are integer multiples of the 1920×1080 (HDTV) picture size. Temporal resolutions for UHDTV can go up to 120 frames per second (fps) with progressive scanning only. It also allows 10- and 12-bit color depth, while the colorimetry system is wider than the one specified in Recommendation ITU-R BT.709 [2] for HDTV content, and covers 75.8% of the CIE 1931 color space. The chrominance sampling ratios included in BT.2020 are 4:2:0, 4:2:2 and 4:4:4.

Based on BT.2020, which defines the parameters of UHDTV services from the signal perspective, other organizations such as Digital Video Broadcasting (DVB) and European Broadcasting Union (EBU) have been working towards the definition of the parameters needed by applications which make use of UHDTV content. DVB has recently ratified the parameters for the delivery of UHDTV ("UHD-1 phase 1") services (they are published as version 2.1.1 of ETSI TS 101 154 [3]): spatial resolution of 3840×2160, maximum bit-depth of 10 bits, temporal resolution up to 60 fps, and BT.709 colorimetry.

Even with the simplest form of UHDTV content, which only increases the number of pixels compared to HDTV, the volume of data associated with UHDTV content is at least four times that for HDTV content. Therefore, in order to reduce the UHDTV burden on the distribution networks, improved compression techniques should be employed when delivering UHDTV services. As an answer to these needs, the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) have finalized the Version 1 of H.265/high efficiency video coding (HEVC) standard [4] in January 2013. HEVC is the state-of-the-art in video compression and can provide the same perceived video quality as its predecessor H.264/advanced video coding (AVC) [5] at half of the bit-rate [6]. For UHDTV content, the MPEG final verification tests have shown an average bit-rate reduction of up to 60% [7].

Even though improved compression technology is key in enabling the delivery of UHDTV content, it is also equally important to distribute the available bit-budget so that the impact of video coding artifacts is minimized. This is particularly true for UHDTV services given the high expectations of audiences. This paper considers as its application scenario the delivery of nearly live UHDTV video over streaming platforms, such as BBC iPlayer, using the HEVC standard. Accordingly, a given

Manuscript received March 7, 2016; revised September 2, 2016; accepted November 14, 2016. Date of publication; date of current version. Part of the work described in this paper has been conducted within the project COGNITUS. This work was supported by the European Union's Horizon 2020 research and innovation programme under Grant 687605. The guest editor coordinating the review of this paper and approving it for publication was Prof. Alexander Raake.

I. Zupancic and E. Izquierdo are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: i.zupancic@qmul.ac.uk; ebroul.izquierdo@qmul.ac.uk).

M. Naccari and M. Mrak are with the British Broadcasting Corporation, London W12 7SB, U.K. (e-mail: matteo.naccari@bbc.co.uk; marta.mrak@bbc.co.uk).

Digital Object Identifier 10.1109/JSTSP.2016.2634458

amount of latency (e.g. few seconds) in the playout is tolerated as well as some bit-rate fluctuations around the target value with time (e.g. up to 30% above the target rate within a period of 3 seconds). This application scenario can be extended to any practical video coding application under constant bit-rate (CBR) constraints. A major requirement is that coding artifacts such as blocking, blurring, and contouring should be minimized. Moreover, the video quality should stay constant over the time, especially when an intra coded frame is inserted because of a scene change.

Rate control guarantees that the available bit-budget is distributed so that the video quality is maximized. A rate control method aims to optimize the visual quality given the limited bandwidth constraints. Generally speaking, rate control can be divided into two main steps. The first one allocates the right amount of bits to each level of the coding process, i.e. structure of pictures (SOP), frame, macroblock or coding unit (CU) in HEVC. In the second step, the allocated rate is used to derive the amount of compression to be applied over a given part of the video sequence.

Rate control can be performed in single- or multi-pass fashion. Single-pass rate control methods allocate the available rate and tune the encoding based on some *a priori* knowledge on the sequence statistics or data collected over previously encoded frames. Contrarily, multi-pass controllers encode a given video segment multiple times, where the results of one step are then used in the subsequent ones. Single-pass rate control is usually employed in applications with real time or very low latency requirements, such as live broadcasting or production. Conversely, multi-pass rate control is usually employed in near real-time applications with continuous scene changes, such as on-demand services, where additional computational complexity can be tolerated.

This paper proposes a two-pass rate control method for streaming of UHDTV content using Version 1 of the HEVC standard. In the first pass, the algorithm performs a pre-encoding analysis, where a light complexity encoder is used to compress the number of frames associated with one intra period, and then collect information such as bit-rate distribution over different frames. This information is then used to fit and update the models used to decide the quantization steps to be used over different frames and image areas, while performing the actual compression in the second pass. The proposed method achieves improved performance compared to existing approaches, especially at the beginning of each scene. The latency introduced by the proposed rate control method is minimal and mainly associated with the pre-processing stage. The proposed rate control method does not imply any additional constraint on the size of the coded picture buffer (CPB). In fact, once the pre-analysis stage is concluded, the actual encoding can start and bit allocation can be adjusted (e.g. on a frame basis) to meet the CPB size constraints specified by HEVC for a particular level and tier. Overall, the main contributions brought by the paper can be summarized as follows:

1) Use of a low complexity pre-encoding step which provides an accurate estimate of the bit-rate profile spent on different frames.
2) Content adaptive initialization of parameters for the rate-quantization step model, based on the data collected during the pre-encoding step.
3) Automatic derivation of initial quantization step for each sequence based on a simplified encoding method which uses multiple quantization steps within a frame.

The remainder of this paper is organized as follows. An overview of the existing rate control methods with the emphasis on the state-of-the-art methods designed for HEVC is presented in Section II. The proposed two-pass rate control method is described in detail in Section III, while Section IV presents a comprehensive experimental validation of the proposed algorithm. Finally, Section V concludes the paper and points out some future work related to the proposed method.

II. Overview of the Related Background

As stated in the Introduction, rate control is one of the essential tools for any practical video codec and consists of two main steps: rate allocation at different granularity levels (e.g. SOP, frame, and block level) and derivation of coding parameters for a given target rate. Over the years, literature has mainly focused on the second step by proposing different models to express the relationship between coding rate and parameters.

This section provides an overview of the existing rate control methods and is organized into four subsections where the first three review the literature associated with models for coding parameters derivation, multi-pass algorithms, and algorithms devoted to improve the perceived video quality. Finally, the fourth subsection focuses on the efficient method based on $R-\lambda$ model, which serves as a basis for our novel two-pass rate control method.

A. Modeling the Coding Rate and Parameters Relationship

One of the first attempts to model the relationship between coding rate and quantization parameter (QP) dates back to the MPEG-2 Video standard with the rate control method implemented in the Test Model 5 (TM5) reference implementation [8]. In this rate control method, the QP value for each macroblock is calculated adaptively based on target bit-allocation and predicted macroblock spatial activity. The Video Model (VM8) used during the development of the MPEG-4 Part 2 (Visual) standard uses a more accurate model based on a second order rate-distortion (RD) relationship [9]. The reference implementation of the AVC standard (Joint Model, JM) uses a rate control method based on a quadratic rate-quantization (R-Q) relationship [10], which relies on the assumption that the residual information follows a Laplacian distribution [11]. The mean absolute difference (MAD) for the residuals is used to estimate the complexity of basic coding units and corresponding QP. Later on, Kamaci *et al.* [12] showed that a Cauchy distribution is more suitable than Laplacian to represent the residuals, and proposed a frame-level rate control method based on these findings.

Based on the well known quadratic R-Q model, Choi *et al.* proposed a rate control method [13] which was used in early versions of HEVC reference software (HM) implementation

[14]. However, due to the flexible quadtree partitioning used in HEVC, this R-Q model is not sufficiently accurate to quantify the relationship between rate and quantization step. Lee and Kim [15] improved this quadratic R-Q model by proposing a new relationship for inter coding in HEVC based on a mixture of multiple Laplacian distributions. Still targeting the HEVC standard, Lee *et al.* [16] proposed instead a frame-level rate control based on different rate allocation models for the bits spent on texture (e.g. residuals) and non-texture (e.g. motion vectors) data. In this approach, multiple Laplacian distributions are used to model the rate of texture bits, while the rate for non-textured bits is modeled with a linear relationship. All these methods based on the R-Q model assume that quantization parameter is a crucial factor in determining the bit-rate. However, that condition holds only when other coding parameters (e.g. the coding mode) are fixed. Given the RD optimization operated in HM, along with the flexible quadtree partitioning specified in HEVC, this assumption is not necessarily true, as already pointed out [17].

Another group of rate control methods tries to build a relationship between the rate and percentage of zeros in quantized transform coefficients ρ. He *et al.* [18] proposed a ρ-domain model and associated rate control. Based on the estimated RD curves, a rate-shape-smoothing algorithm is proposed to smooth the rate distribution and ensure a consistent picture quality. A quadratic ρ-domain rate model was proposed by Wang *et al.* [19] and used in a hierarchical bit-allocation scheme for rate control in an HEVC codec. The proposed algorithm uses a linear relationship in ρ-domain between the bits associated with texture and the number of non-zero transformed coefficients. The number of non-zero transformed coefficients is then modeled as a quadratic function of quantization step. Rate control algorithms based on the ρ-domain relationship work well in fixed transform size coding schemes. Therefore, in video coding standards such as HEVC which specify variable sizes for transform blocks, the relationship between ρ and rate is not sufficiently accurate.

The relation between Lagrange multiplier λ and coding rate was firstly analyzed by Li *et al.* [17]. They proposed hyperbolic R-λ model which shows a higher correlation when compared with the aforementioned R-Q models. The R-λ model was utilized in the state-of-the-art HEVC rate control method, where the bit-budget is allocated using three different levels of granularity. This rate control method was further improved for intra frames [20] using the sum of absolute transformed differences (SATD) as a complexity measure. SATD for original 8×8 blocks is calculated and used to allocate the bit-budget. Based on the R-λ model, two approaches for improved bit-allocation have been recently proposed. Li *et al.* [21] proposed a method for largest CU (LCU)-level bit-allocation in HEVC rate control. In this approach, the formulation for optimal bit-allocation is established using the Lagrange multiplier, computed by minimizing the distortion under the given bit-rate constraints. Then, recursive Taylor expansion method is used to obtain the approximate closed-form solution for the optimal LCU-level bit-allocation formulation. Wang and Ngan [22] proposed a method which uses the distortion of collocated coding tree units (CTUs) in the previous frame to establish a linear relationship between distortion and λ. Based on this distortion model, a different bit-allocation algorithm in λ-domain is applied.

B. Multi-Pass Rate Control Methods

Although parallel architectures are becoming ubiquitous, not many multi-pass rate control methods have been proposed in the past. In x264 [23], which is one of the most popular AVC software implementations, five different rate control modes are specified. Apart from a two-pass approach, where the target number of bits is predicted based on the frame complexity from full encoding in the first pass, one-pass approaches with fast complexity estimation scheme are also available. In this case, a fast motion estimation (ME) algorithm is performed over a half-resolution version of the frame and SATD of the residuals is used as a complexity measure. After encoding each frame or macroblock, future QPs are updated to compensate for mispredictions in rate using short- and long-term compensation schemes. In the context of HEVC, Wen *et al.* [24] proposed a rate control method based on R-λ model with pre-encoding. In the pre-encoding step, the video sequence is encoded using only 16×16 coding units. Rate for the CUs of size 64×64 is then estimated using the rate associated with 16×16 CUs. R-λ model parameters, as well as weights for bit-allocation of 64×64 coding units, are computed using the data from pre-encoding. They also propose a mechanism for resetting the parameters when a scene change leads the existing model parameters to become obsolete.

Another two-pass rate control method for HEVC was proposed by Wang *et al.* [25] based on the structural similiarity (SSIM) index. Coding statistics are collected during the first pass, which is performed using a constant QP. These statistics are then used during the second pass for SOP level bit-allocation. Furthermore, Laplacian-based rate and perceptual distortion models are established to adaptively derive λ and dynamically allocate bits. Rate control at finer granularity levels is performed in a perceptually uniform space. It should be noted that in this case the computational complexity associated with the first pass can be quite high. Deng *et al.* [26] proposed a multi-pass rate control method based on the SATD of the residuals and pre-encoding. Pre-encoding is performed using multiple QP values and a limited set of depths and PU modes to obtain rate, distortion, and SATD data which is then fitted into the SATD-RD model using the least squares method. Estimated data is then used to set the parameters used in rate control. However, this method may be of limited use in practical applications with low latency requirements, due to the computationally expensive pre-encoding step.

C. Rate Control Methods with Region-Based Bit-Allocation

In addition to general purpose rate control methods, specific region-based rate control methods have been proposed in the context of different video coding standards. Hu *et al.* [27] proposed a region-based rate control method for AVC. In this approach, inter-frame information is utilized to divide each frame into multiple regions based on their RD behavior. Macroblocks with similar characteristics are classified into the same region

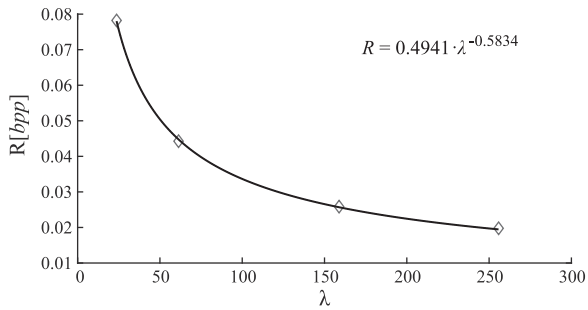

Fig. 1. Fitted R-λ curve for *Manege* test sequence. The sequence was encoded using 4 QP values (27, 31, 35, and 37) and obtained rates are denoted with diamonds.

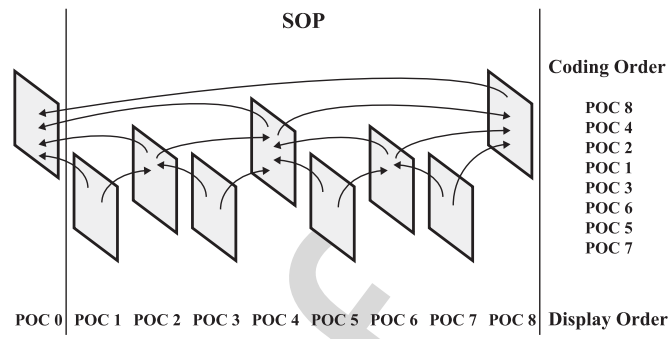

Fig. 2. Random access SOP used in the experiments.

which is treated as a basic unit for the rate control. Recently, Meddeb *et al.* [28] proposed a region of interest (ROI) based rate control method for HEVC. They divide a frame in tiles which correspond to regions with different characteristics. Tiles containing ROIs are then encoded using different encoder settings than non-ROI tiles to achieve better visual quality. The main issue for this kind of methods is the ROI detection which is always content dependent and when erroneously detected, it can lead to poor video quality in regions which attract the attention of the observer.

A method based on perceptual bit-allocation was proposed by Tang *et al.* [29], where a Canny edge detector was used to distinguish between randomly-textured, structurally-textured, and smooth regions. The method allocates fewer bits to randomly-textured regions, given the property of the human visual system which is less sensitive to perceptual distortions in textured image areas. Another bit-allocation method based on a neurobiological model of visual attention was proposed by Lee *et al.* [30], where the model was first used to predict high saliency regions in input frames to generate a saliency map. Based on the human foveated retina characteristic, top salient locations in the saliency map were located and used to generate a guidance map. This guidance map was then used to guide the bit allocation process by tuning the QP values. The approach is based on the study [31] which showed that a saliency map model can accurately predict the human gaze.

D. State-of-the-art HEVC Rate Control Method

It was shown that there exists a robust relation between the rate R (in bits per pixel) and Lagrange multiplier λ which can be expressed with a hyperbolic function [17]:

$$R = a \cdot \lambda^{b}, \tag{1}$$

where a and b are parameters related to the video source. An example of R-λ relationship is shown in Fig. 1. Due to its improved accuracy and robustness, the rate control method based on the R-λ model defined in (1) has been included in the HM reference implementation since version 9.0, and it was there at the time of writing (Version 16.7). The algorithm can be divided into two parts: bit-allocation, and achievement of target bit-rate utilizing the R-λ model. The bit-allocation part is considered at three different levels, namely SOP, frame, and basic unit level. Basic unit in this context is represented by 64×64 CUs, also denoted as CTU in the HEVC standard [4]. When allocating bits at a frame-level, each frame is weighted differently depending on which hierarchical level in the SOP it belongs to, and assuming a random access SOP configuration as used in [32]. A picture structure that corresponds to the SOP configuration used is depicted in Fig. 2, where the picture order count (POC) for each picture is shown to highlight the difference between display and coding order. The random access SOP also defines how the QP changes on a frame basis. More precisely, let QP_{base}, which is an encoding parameter used to generally control the output bit-rate, be the QP value for intra frames, then $QP_{\text{base}} + 1$ will be used for POC 8 frames, $QP_{\text{base}} + 2$ for POC 4 frames, $QP_{\text{base}} + 3$ for POC 2 and POC 6 frames, and $QP_{\text{base}} + 4$ for POC 1, 3, 5 and 7 frames. Throughout this paper, when the QP structure is set according to the aforementioned values, the encoding will be denoted as variable bit-rate (VBR) coding. At basic unit level, the weights to allocate the available bit-budget are calculated dynamically using the prediction error from a collocated basic unit in the previously coded frames belonging to the same temporal layer.

Once the target rate is determined, it is straightforward to determine λ using the inverse of relation (1):

$$\lambda = \alpha \cdot R^{\beta}, \tag{2}$$

where α and β are model parameters. However, the main problem here is how to determine the parameters α and β, which are generally content dependent. Also, in case of random access SOP structure, different temporal layers may have different model parameters, and hence multiple sets of parameters have to be used within the sequence. In the existing approach, the corresponding α and β are continuously updated after encoding one basic unit or one frame. Finally, the QP value is determined as:

$$QP = c_1 \cdot \ln \lambda + c_2, \tag{3}$$

where c_1 and c_2 are set to 4.2005 and 13.7122, respectively. Obviously, QP is rounded to the nearest integer value for practical use. Finally, to keep the video quality consistent, both λ and QP should not change significantly with time. Hence, λ and QP value range is bounded with respect to the values used in previously encoded frame and basic unit.

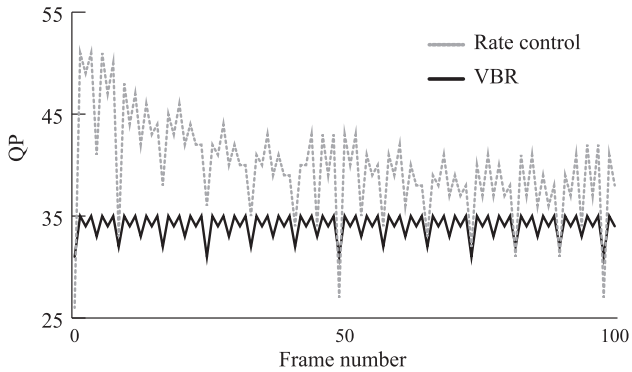

Fig. 3. QP values for the first 100 frames of the *Boxing* test sequence which correspond to the rate obtained with QP value 31 for VBR. QP values used by rate control in HM are denoted with dotted grey line, while QP values associated with the random access SOP are depicted with black.

TABLE I
TEST MATERIAL DESCRIPTION

Sequence name	Fps	Type	Sequence name	Fps	Type
ParkAndBuildings	50	outdoor	TableCar	50	objects
NingyoPompoms	50	objects	TapeBlackRed	60	sport
ShowDrummer1	60	drama	Hurdles	50	sport
Sedof	60	outdoor	LongJump	50	sport
Petitbato	60	outdoor	Discus	50	sport
Manege	60	outdoor	Somersault	50	sport
ParkDancers	50	outdoor	Boxing	50	sport
CandleSmoke	50	drama	Netball	50	sport

Although the rate control method described above shows improved coding performance compared to previous methods proposed for HM, it was noticed that it is significantly under performing at the beginning of the sequence, resulting in degraded quality of experience. In particular, very high QP values (up to 51 in some cases) were used for frames at the beginning of the sequence, as shown in Fig. 3. This is expected, since the initial α and β values for all frame layers are set to predetermined values of 3.2003 and -1.3670, respectively. That is sub-optimal, as λ and corresponding QP value are not calculated using the right model parameters. With model parameters α and β getting continuously updated, the model will gradually become more accurate resulting in better visual quality with time. However, in applications with frequent or continuous scene changes, such as broadcasting, this type of behavior is highly undesirable, as it results in high quality variations of the decoded signal. To overcome this, a two-pass rate control method which accurately predicts parameters α and β, and has small latency is proposed in this paper. In the proposed approach, a short period at the beginning of the sequence is encoded using a reduced set of tools to calculate the initial model parameters which are used to improve the encoding performance, especially at the beginning of the sequence or after a scene change happens.

III. PROPOSED TWO-PASS RATE CONTROL

This section presents the proposed two-pass rate control method for compression of UHDTV video content. Besides describing the proposed method, it is also interesting to analyze the current limitations for the state-of-the-art HEVC rate control method as well as the theoretical performance that can be achieved in case of unlimited computational resources [33], i.e. when the encoder can perform the pre-encoding step testing all possible coding modes to derive the actual bit-rate profile, which is then used in the real encoding step. Throughout the whole section, a fast HEVC encoder implementation based on HM Version 12.0 [14] will be considered and denoted as *HM-fast*. For more details about the *HM-fast* codec, the reader is referred to [34]. The test material and experimental conditions are described in the first subsection. Results and findings from the analysis are reported in the second subsection, while the following subsections describe the proposed method.

A. Test Material and Coding Conditions

The test set used in this paper is composed of 16 sequences with 8 bits per component, 4:2:0 chroma format, 3840×2160 spatial resolution, and frame rate of 50 and 60 fps. The names of these sequences, along with the type of content portrayed are listed in Table I. Each sequence is coded with four QP values. They have been determined by visually inspecting the test set compressed with QP ranging from 22 to 45, to determine a good coverage of different visual quality levels: from very good (i.e. coding artifacts unnoticeable) to fairly poor (i.e. coding artifacts visible and annoying). Content denoted as *outdoor* portrays external scenes. Some of these sequences contain water and complex motion (e.g. *PetitBato*, *Sedof* and *Manege*) or sharp details and camera panning (e.g. *ParkAndBuildings*), and large area picturing grass (e.g. *ParkAndBuildings* and *ParkDancers*). Content denoted as *drama* corresponds to indoor scenes representative of television drama. Content denoted as *objects* represents indoor scene with moving objects. This content is not fully representative of UHDTV material, but given its spatial and temporal features, is challenging from the compression point of view. Finally, content denoted as *sport*, represents various sports content containing indoor and outdoor sequences.

All the sequences have been encoded according to the Joint Collaborative Team On Video Coding (JCT-VC) common test conditions (CTC) [32] using the selected QP values and the random access main (RA-Main) configuration, as this is representative of the encoding settings used in broadcasting services. Throughout this paper, compression efficiency and rate inaccuracy are used as performance metrics. For compression efficiency, the metric used is the Bjøntegaard delta-rate (BD-rate) computed according to [35] between the anchor data (i.e. the sequences compressed with JCT-VC CTC) and the sequences compressed according to the described experiments. In this context, negative BD-rate values will correspond to compression efficiency gains. Given the use of 4:2:0 chroma format, only the BD-rate for the luminance component will be considered. The rate inaccuracy is measured as an absolute percentage deviation from the target rate. Lower value corresponds to higher rate accuracy.

TABLE II
BD-RATE (BD-R) AND RATE CONTROL INACCURACY (I) FOR THE THREE EXPERIMENTS DESCRIBED IN SECTION III-B

	Experiment 1		Experiment 2		Experiment 3	
Sequence	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]
ParkAndBuildings	4.2	1.3	4.7	0.0	2.3	0.0
NingyoPompoms	6.5	0.0	3.5	0.0	3.5	0.0
ShowDrummer1	23.4	0.0	2.2	0.0	1.2	0.0
Sedof	3.9	0.0	5.0	0.0	5.0	0.0
Petitbato	8.8	0.1	2.1	0.1	1.6	0.0
Manege	1.4	0.0	2.6	0.0	2.0	0.0
ParkDancers	5.0	1.1	-0.1	0.0	-0.4	0.0
CandleSmoke	16.2	0.0	4.2	0.0	2.3	0.0
TableCar	8.2	1.8	2.0	0.0	1.3	0.0
TapeBlackRed	13.6	0.2	2.3	0.0	0.8	0.0
Hurdles	8.9	0.1	2.3	0.0	2.3	0.0
LongJump	5.0	0.0	4.0	0.0	3.9	0.0
Discus	4.1	0.0	3.1	0.0	3.8	0.0
Somersault	21.9	0.0	7.5	0.0	1.1	0.0
Boxing	6.5	0.0	1.5	0.0	2.3	0.0
Netball	4.1	0.0	2.4	0.0	1.2	0.0
Average	8.8	0.3	3.1	0.0	2.1	0.0

B. Performance Analysis for the State-of-the-art HEVC Rate Control Method

This section presents the analysis performed over the state-of-the-art HEVC rate control method. Three experiments were conducted. In the first experiment, the coding efficiency of the existing rate control method in HM is measured in terms of BD-rate between the coding performance of the *HM-fast* codec encoded with VBR and *HM-fast* with rate control and using the bit-rate from VBR as target value. The BD-rate and rate inaccuracy for the above described experiment are shown in Table II as *Experiment 1*. As may be noted, the existing rate control method in HM produces significant coding losses compared with VBR encoding. For instance, BD-rate losses larger than 20% are reported in some cases. To investigate the possible source of such high encoding losses, Table III shows the BD-rate measured on different intra periods for every tested sequence. As may be noted, the BD-rate penalty is mostly concentrated at the beginning of the sequence (i.e. in the first intra period). Average BD-rate penalty for the first intra period is considerably higher than in the rest of the sequence. This can be explained by the fact that the existing rate control method in HM uses predetermined parameter values for R-λ model at the beginning of the sequence, since it has no prior knowledge of the content currently being encoded. The rate inaccuracy for all sequences seems to be sufficiently low.

In the second experiment, SOP and frame-level bit-allocation in the HM rate control method were bypassed, and the bit-budget was instead derived from the numbers of bits spent on each frame during VBR encoding. To handle the cases of bit underspending or overspending, a simple rate management scheme was added to redistribute the differential bits to future frames based on their weights associated with the SOP used. The frame weights are determined based on the temporal layer in the SOP a given frame belongs to and their values are reported in [17]. This process is repeated after encoding each frame. The aim of this experiment is twofold: on the one hand, the bit-allocation as designed in the rate control method of HM can be tested and its accuracy assessed. On the other hand, also the accuracy of the R-λ model can also be thoroughly investigated. The BD-rates for the luminance component associated with this experiment are shown in Table II and are denoted as *Experiment 2*. It can be observed that replacing the existing SOP and frame-level bit-allocation, with the frame size obtained from VBR encoding mode, improves the performance significantly. Moreover, the accuracy of achieving the target rate was further improved compared to the existing rate control method in HM.

The third experiment aimed to examine the impact of initializing the model parameters with correct values. As described in SubSection II-D, the initial values of parameters α and β in (2) for all temporal layers in the SOP are set to a predetermined value in the rate control method of HM. In this experiment, the bit-rates and associated λ values, obtained from VBR encoding and using four different QP values, were used to fit the R-λ model from Eq. (2). The fitting is performed differently for each SOP temporal layer and for each sequence. The cost minimized during the fitting is the sum of absolute differences between the QP value predicted by the model and the one used during encoding. The QP derived by the model is obtained by Eq. (3). The reason for minimizing the cost using the QP value is because a poor performance of the rate control method was observed when minimization was applied to λ. In fact, small differences in the λ value may translate into large differences for QP, when λ values are small. The α and β values obtained from the fitting were used to initialize the corresponding parameters for frames of each temporal layer. As in the previous experiment, SOP and frame-level bit-allocation were replaced with the coding rate obtained from VBR encoding. The results of this experiment are shown in Table II as *Experiment 3*. It can be seen that the encoding performance of modified rate control method has been further improved, with rate inaccuracy achieving almost theoretical minimum, i.e. zero.

The results of these experiments show that the existing rate control method in HM can be improved by replacing the SOP and frame-level bit-allocation with the coding rate associated with VBR encoding and initializing the parameters based on fitting the actual rate in the R-λ model. However, in practical applications, this information is not available prior to encoding and in order to obtain it, a full sequence needs to be encoded using at least 3 different QP values, resulting in a massive computational overhead. The proposed rate control method overcomes these complexity issues, as explained in the following subsections.

C. Bit-Rate Profile Analyzer for Pre-Encoding Step

During pre-encoding, a rate control method encodes a given video segment (e.g. one SOP or one intra period) and uses the coding rate to derive the number of bits spent in each frame. Having this information would allow the rate allocation stage to distribute the bit-budget accordingly, where the higher the rate spent on a frame, the higher the bits allocated to it. This

TABLE III
BD-RATE (IN PERCENTAGE) PER INTRA PERIOD (IP) DISTRIBUTION. FOR TEST SEQUENCES WITH 50 FPS IP WAS SET TO 48, WHILE FOR 60 FPS SEQUENCES IP WAS SET TO 64. IN CASE THE SEQUENCE HAS LESS THAN 10 IPS, THE VALUES IN CORRESPONDING FIELDS IN THE TABLE ARE MARKED AS N/A

Sequence	1st IP	2nd IP	3rd IP	4th IP	5th IP	6th IP	7th IP	8th IP	9th IP	10th IP
ParkAndBuildings	6.9	2.7	1.7	3.5	3.8	6.3	7.0	5.2	8.4	6.6
NingyoPompoms	15.5	4.7	5.5	6.6	6.7	4.9	3.4	7.2	4.4	5.6
ShowDrummer1	29.6	9.0	4.5	38.7	29.4	−4.8	−0.9	23.7	N/A	N/A
Sedof	2.9	3.4	3.0	3.5	4.4	4.9	4.4	4.4	3.9	N/A
Petitbato	12.1	4.8	6.2	7.3	11.0	8.3	8.9	10.8	9.7	N/A
Manege	1.4	0.8	0.7	0.9	1.2	1.3	1.8	2.2	2.8	N/A
ParkDancers	4.4	6.6	8.7	6.0	0.7	3.1	14.6	8.8	4.4	1.0
CandleSmoke	32.3	6.2	19.5	7.5	32.3	7.3	22.4	5.7	5.2	5.4
TableCar	5.5	15.0	24.6	−1.8	9.1	0.9	7.9	1.6	2.6	N/A
TapeBlackRed	29.7	4.3	7.0	7.7	7.7	6.9	7.1	5.1	5.7	3.9
Hurdles	9.2	1.7	2.9	3.1	3.3	4.0	2.8	4.2	3.8	12.4
LongJump	5.2	3.3	5.4	3.3	6.4	3.7	11.9	4.7	9.8	2.3
Discus	0.2	5.3	7.9	7.2	5.5	17.9	11.7	N/A	N/A	N/A
Somersault	30.0	13.1	18.6	12.9	7.0	16.0	25.0	10.3	9.6	8.4
Boxing	14.8	4.4	7.8	4.8	4.8	4.6	2.8	2.4	6.8	6.8
Netball	4.3	9.6	4.3	2.1	4.0	2.0	1.2	2.7	4.8	2.7
Average	12.7	5.9	8.0	7.1	8.6	5.5	8.2	6.6	5.8	5.5

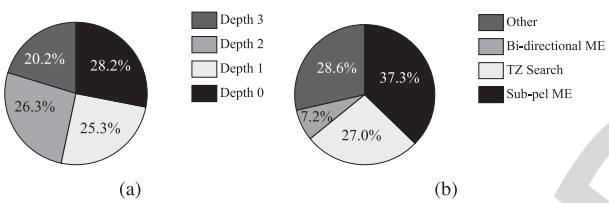

Fig. 4. Percentage of total encoding time spent on testing different coding unit depths (a); and distribution of prediction tasks when the CU depth is equal to zero (b).

TABLE IV
PEARSON CORRELATION COEFFICIENT BETWEEN THE CODING RATE FOR DIFFERENT SOP TEMPORAL LAYERS SPENT BY HM-FAST AND BOTH *SE1* AND *SE2* FOR THE ENTIRE TEST SET

SOP temporal layer	*SE*1	*SE*2
Intra	0.9841	0.9857
0	0.9578	0.9836
1	0.9559	0.9860
2	0.9670	0.9871
3	0.9604	0.9875

pre-encoding step is performed in VBR mode and, ideally, the encoder should test all possible coding modes that would be tested during the actual encoding to obtain a bit-rate profile which is as accurate as possible. However, by doing so, the amount of complexity involved can be prohibitive, even for applications without real time constraints and running on parallel computing architectures. One may be also tempted to re-use the coding modes derived during pre-encoding for actual compression to speed up the whole process. However, given that those modes where derived for a fixed quantization step, i.e. a fixed Lagrange multiplier, they may be sub-optimal when a different QP is selected by the rate control method. Therefore, the coding modes used during pre-encoding can be only partially re-used and the aforementioned claim on computational complexity needs to be carefully addressed.

In the proposed rate control method, a simplified version of *HM-fast* is used. To derive this simplified encoder (SE), the workload associated with *HM-fast* was profiled to identify the most demanding parts in terms of computational complexity. Fig. 4(a) shows the percentage of encoding time spent on testing different CU depths for all sequences belonging to the test material. It can be seen that the most encoding time is spent while testing CUs at depth 0. Hence, testing of depth 0 may be considered as the most important among all the available depths. Fig. 4(b) shows the distribution of prediction tasks for CUs at depth 0 for all sequences belonging to the test set. It can be seen that sub-pel ME is the most time consuming inter-prediction module. That is followed by integer precision ME and bi-prediction. However, it should be noted that some tasks, such as integer precision ME, are critical and cannot be removed without greatly affecting the encoding process.

From this profiling, two configurations for the simplified encoder have been defined and hereafter denoted *SE1* and *SE2*. In *SE1*, the size for each CU is set to 64×64, sub-pel (i.e. half- and quarter-pel) and bi-directional ME are disabled. In *SE2*, 32×32 CUs are also considered, along with half-pel precision ME. Both simplified encoders can significantly reduce the average encoder complexity (by almost 75% for the case of *SE1*), for considerable drop in coding efficiency. However, as stated above, the ultimate goal of the pre-encoding stage is to derive the profile on how the coding rate is spent in relative terms, i.e. what is the percentage of bits spent on a given frame over the total rate used. To measure how accurate the profile derived by both *SE1* and *SE2* is, the Pearson correlation coefficient was measured on a frame basis between the coding rate spent by *HM-fast* and either *SE1* and *SE2*. Table IV shows these correlation coefficients for different SOP layers. As may be noted, even in case of *SE1*, the correlation coefficient is still fairly high. This confirms the validity of using the rate obtained from

TABLE V
PARAMETERS FOR PREDICTING THE RATE FROM SIMPLIFIED ENCODER MODEL FOR DIFFERENT SOP TEMPORAL LAYERS

	SE1		SE2	
SOP temporal layer	k	l	k	l
Intra	0.3986	1.0576	0.4575	1.0493
0	1.1785	0.9722	0.9462	0.9970
1	0.8126	0.9822	0.8535	0.9951
2	1.2695	0.9421	1.4818	0.9492
3	1.8709	0.9011	1.4378	0.9495

simplified encoders to estimate the actual rate in unconstrained VBR mode.

Even though good correlation values are obtained for both encoders, the rate spent by either the simplified encoders (R_{SE}) is on a different scale with respect to the one spent by *HM-fast* (R_{orig}). The reason for this resides in the limited number of coding modes tested by the simplified encoders which results in increased bit-rate compared with encoder operating with the full set of coding tools. To correct the rate values obtained by *SE1* and *SE2*, the following hyperbolic model was used:

$$R_{\text{orig}} = k \cdot R_{SE}^{l}, \tag{4}$$

where k and l are model parameters. It should be noted that different parameter values were used for frames at different temporal layers, as shown in Table V. These parameters were derived by performing the least squares fitting on frame data from the test material. This can be formulated as:

$$\arg\min_{k,l} \sum_{i=0}^{N-1} \left(R_{\text{orig},i} - k \cdot R_{SE,i}^{l}\right)^2, \tag{5}$$

where N is the number of frames from the same SOP temporal layer used for fitting. The output of the pre-encoding stage can be successfully used for SOP and frame-level bit-allocation. However, in order to initialize the parameters for the R-λ model used to derive the QP for each coding block, some additional pre-encoding steps would be required to fit the R-λ curve resulting in increased computational complexity. The next subsection will describe how the proposed rate control method addresses this issue by performing bit-rate profile and model parameters estimation in one pre-encoding step.

D. Pre-Encoding with Variable QP Within Frame

Subsection III-B demonstrated that initializing the R-λ model parameters on a per sequence and QP basis led to improved coding performance of the rate control. However, in practical applications, it is not feasible to encode a sequence with different QP values (e.g. 4 values) in order to fit the R-λ model. This section describes the proposed variable QP (VQP) framework designed to reduce the computational complexity associated with the pre-encoding phase in rate control.

The main idea of VQP framework is to encode different CTUs in a frame with different QP values by performing only one, instead of multiple encodings. Accordingly, different CTUs within a frame are encoded with different QP values which are in relation with λ as described in Eq. (3). The rate obtained for those CTUs is collected separately and used to fit the R-λ model defined in Eq. (2) to obtain parameters α and β.

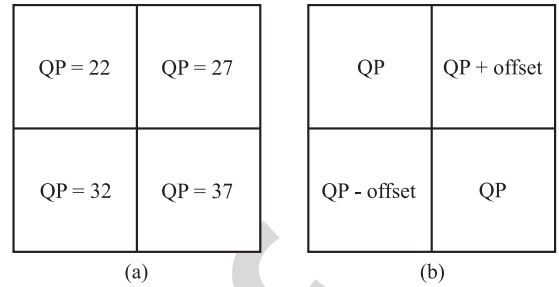

Fig. 5. Variable QP pattern used within a frame for different frame types. Each square represents one CTU. (a) Intra frames. (b) Inter frames.

After the parameters α and β are available, the actual encoding can be performed. It should be noted that the described VQP is not an additional step performed during pre-encoding, but it is a framework applied during the bit-rate profile analysis described in Section III-C. Therefore, no additional processing is required by the proposed VQP.

Besides using VQP to derive the right R-λ model parameters, it should be noted that it can also be used in the decision on the initial QP value for the first intra frame and the pre-encoding stage. In fact, once the R-λ for the video segment under analysis is available, the target rate value is used to derive the associated λ and QP value using Eqs. (2) and (3), respectively. The QP value derived is then used as the value for the bit-rate profile analysis, as well as for the first intra frame.

The main assumption behind the proposed VQP method, is that CTUs sharing the same QP value are representative of the whole statistics associated with the content. To guarantee this, appropriate sampling of the available CUs should be performed. In this paper, two sampling patterns are defined for intra- and inter-coded frames, as depicted in Fig. 5, where each square represents one CTU. Given that the sampling pattern is regular, each QP value will have associated CTUs coming from different image areas. By considering all tested QP values, the derived points on the R-λ model would allow for a more accurate fitting, rather than if the points were derived from CTUs referring to particular image areas (e.g. texture). For intra-coded frames, the four values in Fig. 5(a) are the same as suggested in [32], while in Fig. 5(b) the offset value is set equal to 2. The reason for using two different patterns in intra and inter frames is because R-λ model for intra frames is used to derive the initial QP, so a wider R-λ curve is needed. Therefore, the four QP values as specified in [32] are used. On the other hand, the VQP pattern for inter frames which is used to derive the R-λ model allows statistics to be collected while not interfering significantly with motion estimation and compensation operated by either *SE1* and *SE2*.

E. Workflow of the Proposed Two-Pass Rate Control Algorithm

This section presents the overall workflow associated with the proposed rate control algorithm. As stated above, there are two

Algorithm 1: Processing for the proposed rate control algorithm.

Require: Target bit-rate $\bar{R}$

1: Encode the first frame of the video sequence with the VQP pattern in Fig. 5(a)
2: Collect the coding rate R_{QP} and compute the associated λ for each QP value tested in the VQP pattern
3: Fit the R-λ curve and set the average rate for the first intra picture $\bar{R}_I$ to $\bar{R}/F \times 6$, where F is the frame rate of a sequence
4: Derive the initial QP, QP_{ini} using Eqs. (2) and (3), and $\bar{R}_I$
5: **for all** intra periods in the sequence **do**
6: Encode the current intra period IP with the simplified encoder (*SE1* or *SE2*), encode the intra frame with fixed QP_{ini} and encode the remaining inter frames with the VQP pattern in Fig. 5(b), where QP is determined based on SOP temporal layer of a frame
7: Collect the coding rate R_I for the first intra frame
8: Set $r_2 = \frac{R_{IP}}{R_I}$ as the ratio between the number of bits obtained for the intra period and intra frame
9: Adjust the rate for the intra frame as $R_I \leftarrow R_I \times r_2$ and recompute QP_{ini} using the R-λ curve derived in Step 3
10: For each frame in IP adjust the allocated bit-budget according to the bit-rate profile derived from the simplified encoder
11: Derive parameters α and β for the model in Eq. (2) from the data associated with the tested QP values in the VQP pattern in Fig. 5(b)
12: Run actual encoding using the data for rate control derived in the previous steps
13: **end for**

main processing steps involved: pre-encoding with the proposed VQP method, and encoding with the results gathered from the first step. The processing operated by the proposed rate control method is summarized in the pseudo code of Algorithm 1. Prior to pre-encoding a sequence with VBR mode using VQP framework, the initial QP for the intra frame has to be estimated. Since only the estimated R-λ curve for the intra frame is available prior to performing Step 4 of Algorithm 1, the initial QP is estimated using some previously known statistics. However, when the ratio between the number of bits spent on intra frame and total number of bits spent for all frames in intra period is known (i.e. after completing the pre-encoding for a given intra period), the initial QP value used during the second-pass CBR encoding is recomputed, as described in Step 9.

The overall processing for the proposed rate control method is also depicted in Fig. 6. The pre-encoding stage introduces a delay which can be minimized using multi-threading with one thread dedicated to pre-encoding, so that only one intra period delay (i.e. approximately 1 second) is introduced. It is worth

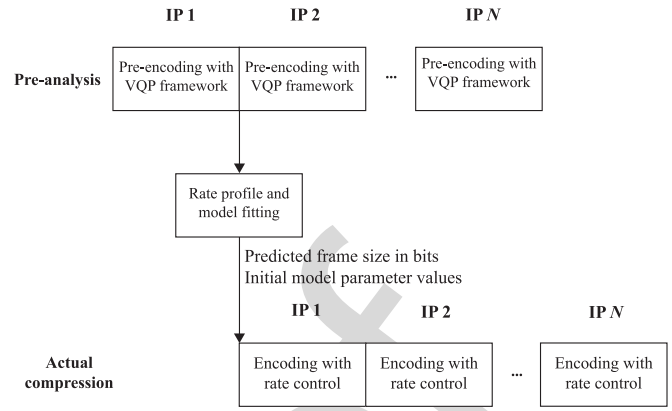

Fig. 6. Block diagram of the proposed approach.

pointing out that the delay resulting from pre-encoding of one intra period does not imply the usage of a CPB of the same size of one intra period. In fact, during the actual encoding (Step 12), the size of the CPB can be set according to the constraints specified in the selected level and tier.

IV. Experimental Results

This section presents the performance of the proposed two-pass rate control method. The test material, coding configuration, and performance indicators are the same as described in Subsection III-A. All the results presented here will use as reference the *HM-fast* codec run in VBR mode. The target rate values fed as input to the proposed rate control algorithm will be therefore the ones associated with *HM-fast* run in VBR. All the tests were run on a Linux cluster of Intel Xeon X3450 with 2.67 GHz clock frequency and 8 GB of RAM.

Table VI shows the experimental results for the proposed two-pass rate control method. When compared to the VBR encoding mode, the proposed rate control method achieves an average BD-rate coding penalty of 2.9% with 14.8% rate inaccuracy. This compares favorably with the state-of-the-art HEVC rate control method which provides on average 8.8% BD-rate losses with 0.3% rate inaccuracy. It should be noted that even though the proposed method provides a lower encoder inaccuracy, it still meets the requirements associated with the application scenario considered in the Introduction (i.e. up to 30% bit-rate deviation from the target value within a period of 3 seconds). It is also interesting to analyze the trade-off between the two defined simplified encoders used in the pre-encoding stage. Therefore, Table VI also shows the BD-rate and rate inaccuracy for *SE1* and *SE2*. As expected, *SE2* provides a better performance, namely in terms of coding efficiency penalty, with respect to *SE1*. When using *SE1* during the pre-encoding and replacing the SOP and frame-level bit-allocation with rate prediction from *SE1*, the proposed rate control method achieves 4.8% BD-rate losses with 15.2% rate inaccuracy. If initial values for α and β parameters are set based on the model fitting using the data obtained from pre-encoding, the proposed encoder achieves on average 3.8% BD-rate losses with 15.2% rate inaccuracy. Even better encoding performance can be obtained if

TABLE VI
EXPERIMENTAL RESULTS IN TERMS OF BD-RATES (BD-R) AND RATE INACCURACY (I). ALL THE TESTS WERE PERFORMED UNDER THE RA-MAIN CONFIGURATION

	HM rate control		SE1 rate control		SE2 rate control		SE1 rate control with param. init.		SE2 rate control with param. init.	
Sequence	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]
ParkAndBuildings	4.2	1.3	6.6	13.7	5.8	15.6	4.7	13.7	4.4	15.6
NingyoPompoms	6.5	0.0	3.9	2.3	3.3	3.1	4.9	2.3	4.3	3.1
ShowDrummer1	23.4	0.0	10.5	36.0	9.8	33.4	9.5	36.0	8.8	33.4
Sedof	3.9	0.0	7.8	11.2	5.8	11.6	6.3	11.2	5.2	11.6
Petitbato	8.8	0.1	-1.2	13.5	-1.2	14.4	-1.0	13.5	-0.7	14.4
Manege	1.4	0.0	10.3	7.0	5.6	7.4	8.7	7.0	4.2	7.4
ParkDancers	5.0	1.1	1.5	5.2	2.4	5.9	1.0	5.2	2.1	5.9
CandleSmoke	16.2	0.0	2.4	13.5	2.5	15.3	0.9	13.5	0.7	15.3
TableCar	8.2	1.8	0.5	4.1	0.7	4.0	-0.2	4.1	-0.9	4.0
TapeBlackRed	13.6	0.2	4.0	4.8	3.3	4.4	2.8	4.8	2.4	4.4
Hurdles	8.9	0.1	5.6	20.3	2.6	19.6	5.3	20.3	2.4	19.6
LongJump	5.0	0.0	6.0	15.2	5.4	14.5	5.1	15.2	3.8	14.5
Discus	4.1	0.0	8.2	77.1	5.2	67.6	5.8	77.1	3.8	67.6
Somersault	21.9	0.0	5.9	5.3	5.4	4.4	1.6	5.3	1.3	4.4
Boxing	6.5	0.0	2.0	4.0	1.6	7.1	2.7	4.0	2.5	7.1
Netball	4.1	0.0	2.7	9.7	2.5	8.1	2.4	9.7	2.2	8.1
Average	8.8	0.3	4.8	15.2	3.8	14.8	3.8	15.2	2.9	14.8

TABLE VII
EXPERIMENTAL RESULTS IN TERMS OF BD-RATES (BD-R) AND RATE INACCURACY (I) FOR THE MODIFIED RATE ALLOCATION PART. ALL THE TESTS WERE PERFORMED UNDER THE RA-MAIN CONFIGURATION

	HM rate control		Modified RC based on SE1		Modified RC based on SE2		SE1 MRC with param. init.		SE2 MRC with param. init.	
Sequence	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]	BD-R [%]	I [%]
ParkAndBuildings	4.2	1.3	6.8	0.3	4.6	0.4	6.5	0.4	5.2	0.4
NingyoPompoms	6.5	0.0	4.1	0.0	5.1	0.0	3.6	0.0	4.5	0.0
ShowDrummer	23.4	0.0	11.4	0.1	9.5	0.1	11.1	0.1	9.0	0.1
Sedof	3.9	0.0	7.7	0.2	5.8	0.2	6.9	0.2	5.0	0.2
Petitbato	8.8	0.1	-0.4	0.0	-1.2	0.1	0.0	0.0	-1.0	0.1
Manege	1.4	0.0	11.0	0.0	9.2	0.0	7.2	0.0	5.3	0.0
ParkDancers	5.0	1.1	2.5	1.3	2.4	1.7	3.2	1.6	3.7	2.3
CandleSmoke	16.2	0.0	8.4	0.5	6.9	0.4	10.5	0.6	7.9	0.5
TableCar	8.2	1.8	1.9	1.1	-0.1	1.4	1.8	1.4	1.1	1.5
TapeBlackRed	13.6	0.2	4.7	0.5	4.2	0.3	4.7	0.6	4.5	0.4
Hurdles	8.9	0.1	8.6	0.0	9.0	0.0	6.0	0.0	5.8	0.0
LongJump	5.0	0.0	5.2	0.0	3.8	0.0	5.9	0.0	3.3	0.0
Discus	4.1	0.0	2.1	0.0	1.5	0.0	2.5	0.0	1.0	0.0
Somersault	21.9	0.0	8.8	0.0	4.4	0.0	9.4	0.0	5.2	0.0
Boxing	6.5	0.0	2.7	0.0	3.2	0.0	2.3	0.0	3.8	0.0
Netball	4.1	0.0	2.2	0.0	2.1	0.0	2.1	0.0	2.0	0.0
Average	8.8	0.3	5.5	0.3	4.4	0.3	5.2	0.3	4.1	0.3

SE2 is used during the pre-encoding. When replacing the SOP and frame-level bit-allocation with rate prediction from *SE2*, the proposed rate control method achieves 3.8% BD-rate losses with 14.8% rate inaccuracy. Finally, when initializing the parameters α and β with data obtained from model using information from *SE2*, 2.9% BD-rate losses can be achieved for 14.8% rate inaccuracy.

To further improve the accuracy of the proposed algorithm, an additional experiment was conducted, whereby the frame-level bit allocation was modified as follows. The weight used to determine the bit budget for each frame was computed as the ratio between the coding bits used for that frame and the total bits spent over the entire intra period by the selected simplified encoder (i.e. *SE1* or *SE2*). The frame weights were then used to allocate the bits at frame level, assuming equal rate distribution among intra periods in the sequence. Table VII shows the associated experimental results. It can be seen that significant accuracy improvements are brought by this new frame-level bit-allocation. When using *SE1* during the pre-encoding and replacing the SOP and frame-level bit allocation with the aforementioned approach, the modified rate control method achieves 5.5% BD-rate losses with significantly reduced rate inaccuracy of 0.3%. If the initial values for parameters α and β are set based on the model fitting using the data obtained from pre-encoding, the modified encoder would achieve an average 4.4% BD-rate losses with 0.3% rate inaccuracy. When considering *SE2*, the modified rate control method achieves on average 5.2% BD-rate losses with reduced rate inaccuracy of 0.3%. Finally, when also initializing the parameters α and β with data obtained from model using information

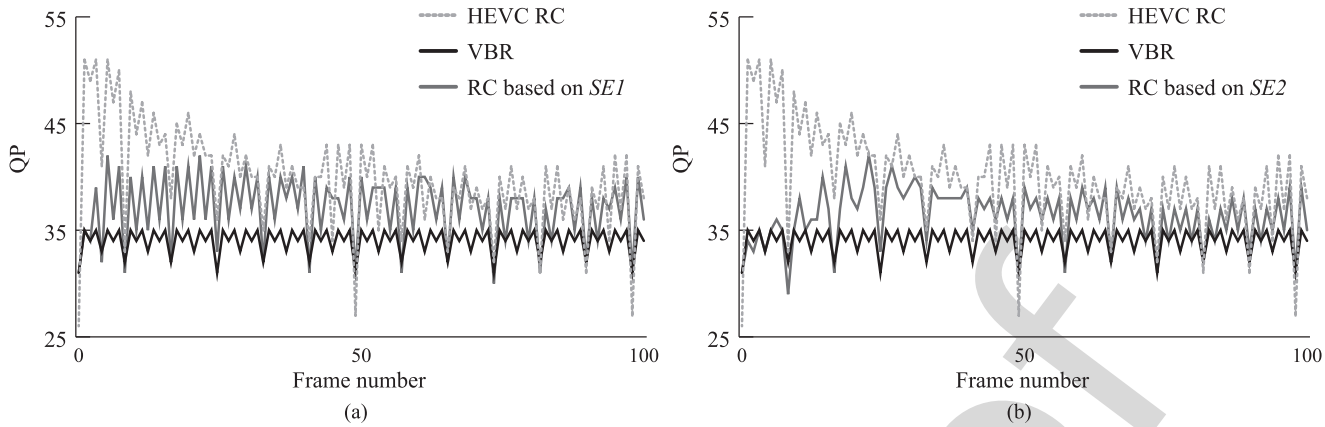

Fig. 7. QP values for the first 100 frames of the *Boxing* test sequence which correspond to the rate obtained with QP value 31 for VBR. QP values for the VBR configuration are depicted with black, QP values used by the state-of-the-art HEVC rate control are denoted with dotted grey line, while QP values used by the proposed rate control method are denoted with solid grey line. (a) Rate control based on *SE1*. (b) Rate control based on *SE2*.

TABLE VIII
STANDARD DEVIATION OF FRAME PSNRS FOR DIFFERENT ENCODERS

	Standard deviation of frame PSNRs			
Sequence	VBR	RC in HM	RC-SE1	RC-SE2
Boxing (QP 23)	0.5269	0.7226	0.4821	0.4891
Boxing (QP 31)	0.5538	0.8956	0.6318	0.5546
ShowDrummer1 (QP 29)	0.7564	0.8113	0.6761	0.6513
ShowDrummer1 (QP 36)	0.4791	0.8181	0.6185	0.5488
Manege (QP 27)	0.5283	0.6940	0.4945	0.5162
Manege (QP 35)	0.5719	0.7814	0.7014	0.7446
TableCar (QP 24)	0.6502	0.6882	0.6603	0.6594
TableCar (QP 30)	0.2705	0.6020	0.3151	0.3349
Petitbato (QP 25)	0.5693	0.7647	0.4746	0.4711
Petitbato (QP 35)	0.6367	0.7676	0.6345	0.6264
Average	0.5543	0.7546	0.5689	0.5596

TABLE IX
AVERAGE SSIM VALUES FOR THE ANCHOR AND DIFFERENT RATE CONTROL METHODS

Sequence	VBR	HM RC	RC SE1	RC SE2	MRC SE1	MRC SE2
ParkAndBuildings	0.963	0.964	0.994	0.994	0.962	0.994
NingyoPompoms	0.968	0.968	0.997	0.997	0.968	0.997
ShowDrummer1	0.860	0.859	0.977	0.977	0.860	0.978
Sedof	0.901	0.901	0.988	0.988	0.899	0.987
Petitbato	0.840	0.838	0.946	0.946	0.840	0.948
Manege	0.876	0.877	0.974	0.975	0.869	0.976
ParkDancers	0.866	0.868	0.965	0.965	0.867	0.965
CandleSmoke	0.897	0.897	0.985	0.985	0.897	0.984
TableCar	0.862	0.864	0.984	0.984	0.864	0.985
TapeBlackRed	0.969	0.968	0.986	0.986	0.968	0.986
Hurdles	0.950	0.950	0.984	0.984	0.949	0.985
LongJump	0.951	0.950	0.989	0.990	0.950	0.989
Discus	0.942	0.935	0.963	0.966	0.936	0.975
Somersault	0.950	0.950	0.977	0.977	0.950	0.977
Boxing	0.959	0.959	0.998	0.998	0.959	0.997
Netball	0.952	0.952	0.984	0.984	0.951	0.983
Average	0.919	0.919	0.981	0.981	0.918	0.982

from *SE2*, 4.1% BD-rate losses can be achieved for 0.3% rate inaccuracy.

It should be noted that the complexity of the second pass of the proposed rate control method is not different with respect to the one of the HM rate control method. As illustrated in Fig. 6, pre-encoding stage introduces a small latency required to process the frames related with the first intra period. Using parallel processing would limit the latency to only initial pre-encoding for the first intra period.

The version of the HM codec used in the experiments does not implement any scene change detector. However, the behaviour of the proposed rate control at the beginning of a sequence is equivalent to what happens after a scene change. In fact, when a scene change happens, the parameters α and β of the R-λ model will be reset to their initial values. Moreover, the internal buffers used to keep track of the QP and λ values for clipping purposes will be also emptied. This resembles to the same initial condition at the beginning of the sequence.

As described in Subsection II-D, when using the existing rate control method in HM, QP values at the beginning of the sequence tend to be much higher than in the VBR case, resulting in degraded Quality of Experience. Fig. 7 shows the comparison of QP values used at the beginning of the sequence between the existing and the proposed rate control method. It can be seen that QP values used by the proposed method are considerably lower than those used by the existing method, and generally correlate more with QP values from the VBR encoding mode.

Furthermore, since one of the aims of the rate control is to smooth the visual quality fluctuations in time, visual quality can also be quantified as the standard deviation of frame-based PSNR values. Table VIII shows the standard deviation of frame PSNRs for some of the sequences from the test set. It can be seen that the standard deviation of PSNR values obtained for the proposed rate control method based on *SE1* and *SE2* are significantly lower than the one associated with the HEVC rate control method. Furthermore, the standard deviation of PSNR values obtained for the rate control method based on both *SE1* and *SE2* are very close to the one of unconstrained VBR encoding mode.

Finally, the perceptual SSIM metric was computed for the anchor and the proposed rate control method with different bit allocation schemes. The average SSIM values for test sequences are shown in Table IX. It can be seen that almost all versions of the proposed rate control algorithm achieve considerably higher perceptual quality when compared with the rate control method in HM. This verifies the claim that the proposed rate control methods also improve perceptual quality.

V. Conclusion

UHDTV is expected to deliver an enhanced visual quality TV services with the improved Quality of Experience compared to the existing HDTV services. Apart from higher spatial resolution, UHDTV has a potential to deliver wider color gamut, high dynamic range and high frame rate. To allow for more efficient delivery of such an enormous amount of data, the current state-of-the-art HEVC standard has been recently developed and standardized. It greatly outperforms the previous video coding standards in terms of compression efficiency. However, when transmitting a video sequence over a limited bandwidth network, visual quality fluctuation with time plays a crucial role to provide the high Quality of Experience. Rate control in video coding aims to optimize the bit-distribution to achieve the highest possible video quality for a given bandwidth constraint. However, in many practical applications with frequent scene changes, the existing rate control methods perform sub-optimal, resulting in degraded visual quality at the scene beginning. To overcome this issue, a two-pass rate control method was proposed in this paper. A simplified encoder was used in the pre-encoding stage to obtain the bit-rate profile for each intra period. A variable QP framework was designed to avoid encoding a sequence multiple times for tuning the model parameters. When compared with VBR encoding mode, the proposed two-pass rate control method achieves on average lower compression losses, 2.9% BD-rate losses compared to 8.8% BD-rate losses for the state-of-the-art HEVC rate control method. The proposed method also achieves significantly higher visual quality. Future research on the proposed method may involve integration of the hypothetical reference decoder (HRD) model in the rate allocation process with variable buffer size, as well as the use of perceptual models to distribute the available bit-budget within one picture to further improve the perceived video quality.

References

[1] ITU-R, "Parameter values for ultra-high definition television systems for production and international programme exchange," ITU-R, Geneva, Switzerland, Tech. Rep. BT.2020, Aug. 2012.

[2] ITU-R, "Parameter values for the HDTV standards for production and international programme exchange," ITU-R, Geneva, Switzerland, Tech. Rep. BT.709-6, Jun. 2015.

[3] ETSI, "Digital video broadcasting (DVB); Specification for the use of video and audio coding in broadcasting applications based on the MPEG-2 transport stream," ETSI, Sophia-Antipolis, France, Tech. Rep. TS 101 154, Mar. 2015.

[4] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[5] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[6] J. Ohm, G. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—Including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[7] T. K. Tan *et al.*, "Video quality evaluation methodology and verification testing of HEVC compression performance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 76–90, Jan. 2016.

[8] Test Model 5, Feb. 2016. [Online]. Available: http://www.mpeg.org/MPEG/MSSG/tm5/

[9] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[10] K.-P. Lim, G. Sullivan, and T. Wiegand, "Text description of joint model reference encoding methods and decoding concealment methods," Joint Video Team, Tech. Rep. JVT-N046, Jan. 2005. Q3

[11] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[12] N. Kamaci, Y. Altunbasak, and R. Mersereau, "Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.

[13] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajic, "Rate control based on unified RQ model for HEVC," Tech. Rep. JCTVC-H0213, Feb. 2012. Q4

[14] HM Reference Software, Jan. 2016. [Online]. Available: https://hevc.hhi.fraunhofer.de/HM-doc/

[15] B. Lee and M. Kim, "Modeling rates and distortions based on a mixture of Laplacian distributions for inter-predicted residues in quadtree coding of HEVC," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 571–574, Oct. 2011.

[16] B. Lee, M. Kim, and T. Nguyen, "A frame-level rate control scheme based on texture and nontexture rate models for high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 465–479, Mar. 2014.

[17] B. Li, H. Li, L. Li, and J. Zhang, "Lambda domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, Sep. 2014.

[18] Z. He, Y. K. Kim, and S. Mitra, "Low-delay rate control for DCT video coding via rho-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[19] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Quadratic rho-domain based rate control algorithm for HEVC," in *Proc. 2013 IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 1695–1699.

[20] M. Karczewicz and X. Wang, "Intra frame rate control based on SATD," Tech. Rep. JCTVC-M0257, Apr. 2013.

[21] S. Li, M. Xu, and Z. Wang, "A novel method on optimal bit allocation at LCU level for rate control in HEVC," in *Proc. 2015 IEEE Int. Conf. Multimedia Expo*, Jun. 2015, pp. 1–6.

[22] M. Wang and K. N. Ngan, "Optimal bit allocation in HEVC for real-time video communications," in *Proc. 2015 IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 2665–2669.

[23] L. Merritt, "x264: A high performance H.264/AVC encoder," Feb. 2016. [Online]. Available: http://akuvian.org/src/x264/overview_x264_v8_5.pdf

[24] J. Wen, M. Fang, M. Tang, and K. Wu, "R-lambda model based improved rate control for HEVC with pre-encoding," in *Proc. 2015 Data Compression Conf.*, Apr. 2015, pp. 53–62.

[25] S. Wang, A. Rehman, K. Zeng, and Z. Wang, "SSIM-inspired two-pass rate control for high efficiency video coding," in *Proc. 2015 IEEE 17th Int. Workshop Multimedia Signal Process.*, Oct. 2015, pp. 1–5.

[26] L. Deng, F. Pu, S. Hu, and C.-C. J. Kuo, "HEVC encoder optimization based on a new RD model and pre-enoding," in *Proc. 2013 Picture Coding Symp.*, Dec. 2013. Q5

[27] H.-M. Hu, B. Li, W. Lin, W. Li, and M.-T. Sun, "Region-based rate control for H.264/AVC for low bit-rate applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 1564–1576, Nov. 2012.

[28] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "ROI-based rate control using tiles for an HEVC encoded video stream over a lossy network," in *Proc. 2015 IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 1389–1393.

[29] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual sensitivity guided bit allocation for video coding," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 11–18, Feb. 2006.

[30] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image Vision Comput.*, vol. 29, no. 1, pp. 1–14, Jan. 2011.

[31] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[32] F. Bossen, "Common test conditions and software reference configurations," Tech. Rep. JCTVC-L1100, Oct. 2012.

[33] I. Zupancic, M. Naccari, M. Mrak, and E. Izquierdo, "Studying rate control methods for UHDTV delivery using HEVC," in *Proc. 2016 Int. Symp. ELMAR*, 2016, pp. 47–50.

[34] M. Naccari, A. Gabriellini, M. Mrak, S. Blasi, I. Zupancic, and E. Izquierdo, "HEVC coding optimisation for ultra high definition television services," in *Proc. 2015 Picture Coding Symp.*, May 2015, pp. 20–24.

[35] G. Bjøntegaard, "Improvements of the BD-PSNR model," ITU-T SG16/Q6, 35th VCEG Meeting, Doc.VCEG-AI11, Jul. 2008.

Ivan Zupancic (S'15) received the B.Sc. degree in computing and the M.Sc. degree (*cum laude*) in information and communication technologies from the Faculty of Electronic Engineering and Computing, University of Zagreb, Zagreb, Croatia, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree with the Multimedia and Vision Research Group, Queen Mary University of London, London, U.K. His research interests include image and video coding.

Matteo Naccari (M'08) received the Laurea degree in computer engineering and the Ph.D. degree in electrical engineering and computer science from the Technical University of Milan, Milan, Italy, in 2005 and 2009, respectively. After earning the Ph.D. degree he spent more than two years as a Postdoc at the Instituto de Telecomunicações, Multimedia Signal Processing Group, Lisbon, Portugal. Since September 2011, he joined BBC R&D, London, U.K., as a Senior Research Engineer working in the video compression team. His research interests are mainly focused in the video coding area where he works or has worked on video transcoding architectures, error resilient video coding, automatic quality monitoring in video content delivery, subjective assessment of video transmitted through noisy channels, integration of human visual system models in video coding architectures and encoding techniques to deliver ultra high definition content in broadcasting applications. He also actively participates to the standardization activities led by the Joint Collaborative Team On Video Coding where he has served as a Co-editor for the specification text of the HEVC Range Extensions. He has authored more than 30 scientific publications for journals, conferences, and book chapters.

Marta Mrak (SM'13) received the Dipl.-Ing. and M.Sc. degrees in electronics engineering from the University of Zagreb, Croatia, and the Ph.D. degree from the Queen Mary University of London, London, U.K. She was a Postdoctoral Researcher with the University of Surrey, U.K. and the Queen Mary University of London. She joined the BBC's Research and Development Department in 2010, to work on video compression research and the H.265/High Efficiency Video Coding standardization. She has co-authored over 100 papers and co-edited two books. She is an elected Member of IEEE Signal Processing Society technical committee on Multimedia Signal Processing and IEEE Circuits and Systems Society technical committee on Multimedia Systems and Applications. She is also acting as an Area Editor of *Signal Processing Image Communication* journal. She has been involved in several projects funded by the European Commission and U.K. research and innovation funds (EPSRC and InnovateUK), and she is coordinating Horizon 2020 Innovation Action COGNITUS, while also acting as a Supervisor in Marie Curie ITN ProVision.

Ebroul Izquierdo (M'97–SM'03) received the M.Sc., C.Eng., and Ph.D. degrees in 1990 and 1993, respectively. Q6

He is a Chair of Multimedia and Computer Vision and Head of the Multimedia and Vision Group in the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K. He holds several patents in the area of multimedia signal processing, and has authored or co-authored more than 500 technical papers including books and chapters in books.

Prof. Izquierdo is a Fellow of the Institution of Engineering and Technology and a Member of the British Machine Vision Association. He is a Chartered Engineer, Member of the Visual Signal Processing and Communications Technical Committee of the IEEE Circuits and Systems Society and a Member of the Multimedia Signal Processing Technical Committee of the IEEE. He is an IEEE Circuits and Systems Society Distinguished Lecturer from 2016 to 2017. He has been an Associated Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2002–2010) and the IEEE TRANSACTIONS ON MULTIMEDIA (2010–2015). He is a Member of the editorial board of several other journals in the field. He has been a Member of the Organizing Committee of several conferences and workshops in the field of image and video processing including the IEEE International Conference on Image Processing, the IEEE International Conference on Acoustics, Speech, and Signal Processing, the IEEE International Symposium on Circuits and Systems, the IEEE Visual Communications and Image Processing Conference, and the IEEE International Conference on Multimedia and Expo.