SSIM-Based Fuzzy Video Rate Controller for Variable Bit Rate Applications of Scalable HEVC

Farhad Raufmehr
Department of Communications Engineering, University of Sistan and Baluchestan, Zahedan, Iran
farhad.raufmehr@gmail.com
Mehdi Rezaei*
Department of Communications Engineering, University of Sistan and Baluchestan, Zahedan, Iran
mehdi.rezaei@ece.usb.ac.ir

Received: 24/May/2019 Revised: 24/Aug/2019 Accepted: 20/Nov/2019

Abstract
Scalable High Efficiency Video Coding (SHVC) is the scalable extension of the latest video coding standard H.265/HEVC. Video rate control algorithm is out of the scope of video coding standards. Appropriate rate control algorithms are designed for various applications to overcome practical constraints such as bandwidth and buffering constraints. In most of the scalable video applications, such as video on demand (VoD) and broadcasting applications, encoded bitstreams with variable bit rates are preferred to bitstreams with constant bit rates. In variable bit rate (VBR) applications, the tolerable delay is relatively high. Therefore, we utilize a larger buffer to allow more variations in bitrate to provide smooth and high visual quality of output video. In this paper, we propose a fuzzy video rate controller appropriate for VBR applications of SHVC. A fuzzy controller is used for each layer of scalable video to minimize the fluctuation of QP at the frame level while the buffering constraint is obeyed for any number of layers received by a decoder. The proposed rate controller utilizes the well-known structural similarity index (SSIM) as a quality metric to increase the visual quality of the output video. The proposed rate control algorithm is implemented in HEVC reference software and comprehensive experiments are executed to tune the fuzzy controllers and also to evaluate the performance of the algorithm. Experimental results show a high performance for the proposed algorithm in terms of rate control, visual quality, and rate-distortion performance.

Keywords: Fuzzy Control, Quality, Rate, Scalable high-efficiency video coding (SHVC), SSIM, Variable bit rate (VBR).

1- Introduction

Multimedia and video technology improvements resulted in a new video coding standard which is called High Efficiency Video Coding (H.265/HEVC). The first version of HEVC was completed in January 2013. This new video coding standard has about 50% bit-rate reduction compared to the previous standard (H.264/AVC) [1, 2]. But video transmission over various networks usually faces many challenges such as diverse end-users and different connections quality. The solution to this problem is scalable video coding (SVC) [3]. So the need for a scalable video coding standard motivated the joint collaborative team on video coding (JCT-VC) to propose a new scalable extension in the second version of HEVC which was published in January 2015 [4]. This newly published version also includes the 3D/Multi-view and Range extensions [5]. The scalable extension of HEVC which name is Scalable High-efficiency Video Coding (SHVC) not only supports the conventional scalable features such as temporal, spatial, and quality scalabilities but also supports new scalability features such as hybrid codec, bit depth, and color gamut. This new scalable extension was proposed by only modifying the first version of HEVC at high-level syntax and the encoding core is unchanged [4].

The available bandwidth and buffering constraints are other challenges in video transmission. Rate control algorithms (RCA) are utilized to solve this problem. According to the tolerable delay, video transmission applications are divided into constant bit rate (CBR) and variable bit rate (VBR) applications. In CBR applications such as conversational applications, the end-to-end delay is crucial, therefore; a CBR RCA with a relatively small buffer is required. The CBR rate control algorithms try to control the short-term average bit rate strictly in order to prevent the small buffer from overflow and underflow. Strict rate control means high fluctuations in the quantization parameter and thereafter a low-level perceptual quality. In VBR applications such as video streaming and broadcasting a relatively higher delay is tolerable so a larger buffer and a VBR RCA can be used in which a loose control over the long-term average bit rate is imposed. The initial buffering delay in VBR applications is higher than CBR applications [6]. For many VBR

* Corresponding Author:

applications, a higher bit rate is usually used to guarantee acceptable video quality. Also, the rate-distortion performance of VBR is much better than CBR [7]. Considering these practical applications, three encoding configurations, including All Intra, Random Access, and Low Delay configurations, are introduced for HEVC. In these configurations, selected Level and Tier determine the buffering capabilities of the decoder. The standard coded picture buffer (CPB) size, decoded picture buffer (DPB) size, and maximum bit rate are sample parameters related to the buffering capabilities. According to these, the random access (RA) configuration is appropriate for VBR applications. The compression performance of the RA configuration is relatively high, but it includes structural coding delay [6].

1-1- Related Works

There are several RCAs including CBR and VBR algorithms proposed for HEVC and SHVC which are reviewed here. Li et al. proposed a rate-λ (Lagrange multiplier) based RCA for the first version of HEVC which enables the encoder to select among coding parameters to achieve the target rate as well as to minimize distortion [8]. Marzuki et al. take the advantages of the rate-λ model to propose a tile-level rate controller for HEVC on the tile parallelization case [9]. Wang et al. proposed a distortion model, a rate model, and a mixed distribution model for residual signal and developed a ρ-domain RCA [10]. Choi et al. proposed a precise RCA based on a rate-quantization model [11]. Seo et al. considered the bandwidth and buffering constraints and proposed a video quality controller based on a distortion-quantization model and a rate-quantization model [12]. Lee et al. developed a rate-quantization model for each Coding Unit (CU) depth based on texture and non-texture models and proposed a frame-level rate control scheme [13]. Wang et al. improved the performance of the conventional rate-λ model by proposing a gradient-based rate-λ model for intra-frame rate control [14].

All the algorithms mentioned above operate as CBR algorithms which are not suitable for the VBR applications. Lopez et al. proposed a VBR control algorithm based on long-term and short-term sliding windows [15]. Inspiring from the idea proposed by Rezaei et al. in [16] as the semi-fuzzy rate controller, Fani et al. and Kamran et al. proposed fuzzy rate controllers for GOP-level and frame-level, respectively [17, 18]. Fani et al. utilized the proportional, integral and derivative components of the GOP bit error as the inputs of a fuzzy system to propose a novel PID-fuzzy video rate controller [19]. Although these RCAs are targeted for VBR applications they have been designed for the non-scalable version of HEVC and it is essential to design appropriate ones for the SHVC.

According to [20, 21], Li et al. extended the idea of the rate-λ model-based rate controller of HEVC to SHVC. Biatek et al. proposed an adaptive rate controller for SHVC which dynamically adjusts the bit rate ratio between the base layer (BL) and an enhancement layer (EL) to optimize the coding performance under the global bitrate constraint [22]. However, these two algorithms fall into the CBR category too. Considering high delay applications of SHVC we proposed a fuzzy-logic-based scalable video rate controller, which falls into VBR [23].

The distortion models which are used in most of the previously discussed RCAs are based on the error-sensitive metrics such as mean square error (MSE) and peak signal to noise ratio (PSNR). The simplicity of calculation is the main popularity reason for these metrics. These metrics are purely mathematical and they do not consider the characteristics of the human visual system (HVS) so they have less correlation with HVS. Since the video quality is ultimately judged by human eyes it is better to utilize metrics with more adaptation to characteristics of HVS. The HVS reacts quickly to the structural information in the field of viewing, so it is better to use structural similarity-based metrics such as the well-known structural similarity index (SSIM) which exploits structural information to estimate the quality of a compressed video [24]. There are several RCAs that utilized the SSIM as a distortion metric in order to increase perceptual video quality. Zhao et al. incorporated the SSIM into the HEVC rate-distortion optimization (RDO) framework and a CU-level RCA [25]. Zeng et al. utilized the SSIM in the video quality assessment and bit allocation scheme and improved the rate-λ model to control the bit rate [26]. Gao et al. considered the bit allocation as a resource allocation problem and by defining an SSIM-based utility function proposed a Nash bargaining solution [27]. These algorithms are CBR but Wang et al. proposed an SSIM-motivated two-pass VBR rate controller for HEVC in which collected information during the first pass is used for bit allocation and bit rate control in the second pass [28]. Zupancic et al. proposed a two-pass rate controller that occupies a fast encoder in the first pass to collect necessary information for rate allocation and model parameter estimation and use the collected information in the second pass [29].

In this paper, inspiring from the presented algorithm in [23], we propose an SSIM-based fuzzy video rate controller for VBR applications of the SHVC which is able to improve the SSIM-based quality measures for compressed video. It controls the bit rate of several temporal, spatial, and quality layers at the same time. The proposed algorithm tries to achieve long-term average target rates for the SHVC video layers by smooth changing of QP used for encoding each layer. In the proposed algorithm, a fuzzy controller is used for each layer to minimize the changes of QP at the frame level while the

buffering constraints are obeyed. Moreover, an SSIM-based quality controller in cooperation with the fuzzy controller is used in each layer in order to improve and smooth the SSIM metric over encoded video frames. The SSIM-based quality controller suppresses the unnecessary QP fluctuation allowing more fluctuation in bit rate and buffer occupancy. The proposed RCA provides encoded videos with smooth and high visual quality. All conventional rate controllers use a target bit rate as the main reference point in the control process and therefore the bit rate is pushed toward a constant value which is unwelcoming for the VBR applications. In our proposed algorithm, in fact, the QP and SSIM are used as references and the attempt is to prevent unnecessary changes of QP and SSIM. This leads to controlled variations in bit rate and smooth visual quality of the compressed video. Using these references enables the rate controller to operate in a wide rate-distortion (R-D) range, which is the main characteristic of VBR rate controllers.

The rest of this paper is organized as follows. First, the details of our proposed RCA are explained in Section 2. Then, some experimental results are reported in Section 3. Finally, conclusions are given in Section 4.

2- Proposed Rate Control Algorithm

The block diagram of our proposed RCA is shown in Fig. 1. A fuzzy rate controller, an SSIM-based quality controller, a number of virtual buffers, and a number of multiplexers are the main parts of the diagram. The algorithm operates at GOP (groups of pictures) level. It computes a base QP for each GOP and the well-known QP cascading technique is used to calculate a QP for each frame in the GOP. In the base QP calculation process, we consider the correlation between coding complexities of consequent GOPs in a scene. So the coding complexity of previous GOP is used as an estimate for that of the current GOP. Therefore, the base QP of previous encoded GOP is used as an estimate for that of the current GOP and then the fuzzy rate controller and the SSIM-based quality controller adjust the base QP by:

$$BaseQP_b^d = BaseQP_{b-1}^d + \Delta QPF_b^d + \Delta QPQ_b^d \ , \qquad (1)$$

where $BaseQP_b^d$ is the calculated QP for the b^{th} (current) GOP. ΔQPF_b^d and ΔQPQ_b^d denote the base QP changes calculated by the fuzzy rate controller and the SSIM-based quality controller, respectively. b and d denote the indices of GOP and layer, respectively. In fact, we can say the base QP of current GOP, consist of the delayed version of the base QP used for previous GOP plus the base QP changes that are calculated by the fuzzy rate controller and the SSIM-based quality controller. The details of the proposed RCA are discussed in the following subsections.

2-1- Virtual Buffer

As shown in Fig. 1 we employed a virtual buffer denoted by ($Buffer^d$) for each layer in order to simulate the buffering process at the decoder side. The buffer size (BS^d) and the target rate (TR^d) are determined by the users according to the bandwidth, buffering, and delay constraints.

Inspiring from the fact that the sub-stream of each layer is multiplexed with those of other layers in the scalable video encoder in order to produce the scalable bitstream, so in the buffering process, the consumed bits of each layer are aggregated with those of lower layers. This is done by layer multiplexers (MUX^d) which is used to simulate the encoder multiplexing process as:

$$MB_a^d = \sum_{j=0}^{d} B_a^j \ , \qquad (2)$$

where B_a^j denotes the consumed bits for the a^{th} frame in the j^{th} layer and MB_a^d is the multiplexed consumed bits for the a^{th} frame in the d^{th} layer. Then, the output of the d^{th} multiplexer is used to update the occupancy of the d^{th} virtual buffer after encoding a^{th} frame according to (3):

$$BO_a^d = BO_{a-1}^d - MB_a^d + \frac{1}{F}\sum_{j=0}^{d} TR^j \ , \qquad (3)$$

Here, BO_a^d denotes the buffer occupancy of the d^{th} layer after encoding the a^{th} frame. Also TR^j is the target rate of the j^{th} layer and F stands for the frame rate. It is notable that we assume %60 of the virtual buffer size (BS^d) is initially occupied by initial buffering i.e.

$$BO_0^d = 0.6 \times BS^d \ , \qquad (4)$$

2-2- Fuzzy Rate Controller

As a conventional fuzzy controller, our fuzzy rate controller contains a fuzzifier, a defuzzifier, a fuzzy interface engine, and a fuzzy rule base. The fuzzifier maps the crisp inputs to the input fuzzy sets. The fuzzy interface engine maps the input fuzzy set to the output fuzzy set according to the fuzzy rule base and the defuzzifier maps the output fuzzy set into a crisp output. The fuzzy rule base is presented with linguistic variables to appropriately link daily conversations to a mathematical framework [30]. We choose the fuzzy logic controller because many non-linear relations that exist in video rate control can be easily included in the fuzzy rules and membership functions (MSF).

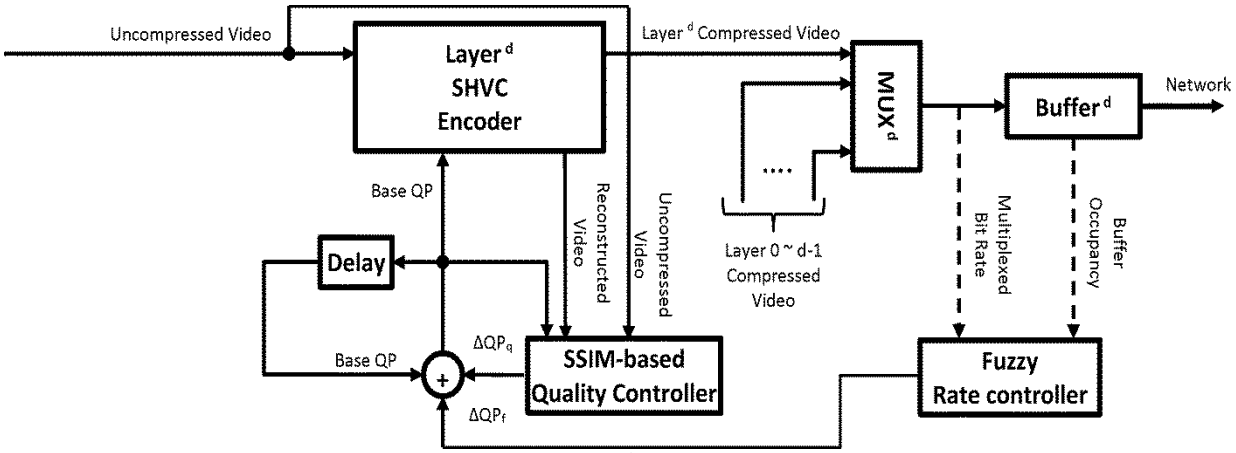

Fig. 1 Block Diagram of Proposed Algorithm

As shown in Fig. 1, the fuzzy rate controller uses two feedback signals from each layer to make an output for each layer. As previously discussed, according to the bandwidth, buffering, and delay constraints, the target rate (TR^d) and the virtual buffer size for each layer are determined by the user. Then, the target rate of each layer would be aggregated with those of all prior layers to make a target rate for the output of each multiplexer.

After encoding a complete GOP with layers, the multiplexed consumed bits and the buffer occupancy are normalized by the aggregated target rate and the buffer size, respectively in order to form two inputs from each layer to the fuzzy rate controller as (5) and (6):

$$x_1^d = \frac{BO^d}{BS^d}, \tag{5}$$

$$x_2^d = \frac{PGC^d}{TGC^d}, \tag{6}$$

where x_1^d and x_2^d are two fuzzy inputs from the d^{th} layer. PGC^d denotes the multiplexed consumed bits by previous GOP and TGC^d stands for the target multiplexed bits for the GOP. The target multiplexed bits for a GOP is calculated by as:

$$TGC^d = N_{GOP}^d \times \frac{1}{F}\sum_{j=0}^{d} TR^j, \tag{7}$$

where N_{GOP}^d denotes the number of frames in a GOP of the d^{th} layer. Taking the expert experiences into account, we design 9 and 7 trapezoidal membership functions (MSF) for x_1^d and x_2^d, respectively as shown in Fig. 2.

We utilized the trapezoidal MSFs because, the non-linear relationships are better included in them than the triangular ones and also they are less computational complex than the Gaussian MSFs. The rule base of our fuzzy system is summarized in Table 1. In this table, the letters L, M, H stand for low, medium and high respectively as the primary term of linguistic variables. Moreover, the letters A, U, E and V stand for absolute, ultra, extra, and very, respectively for the linguistic hedges. As an example of fuzzy rules, we can say: IF x_1^d is Medium-Low and x_2^d is Medium, THEN the output is Medium-High. Also, the desired central values of the fuzzy output corresponding to Table 1 are obtained and shown in Table 2.

In designing the fuzzy membership functions and the desired central values, the attempt is to minimize the QP fluctuations in order to provide smooth and high visual quality for the compressed video while the buffering constraints are obeyed, so the structure of fuzzy rules and MSFs are designed asymmetrically.

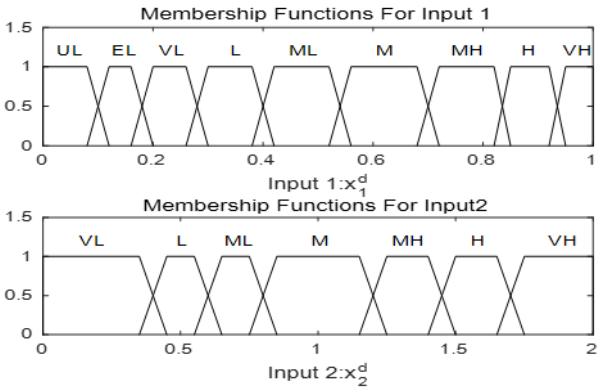

Fig. 2 Membership functions of fuzzy inputs linguistic variables

Table 1 Fuzzy Rule in Linguistic Variables

x_2^d \ x_1^d	UL	EL	VL	L	ML	M	MH	H	VH
VH	AH	AH	AH	UH	EH	VH	H	MH	M
H	AH	AH	UH	EH	VH	H	MH	M	ML
MH	AH	UH	EH	VH	H	MH	M	ML	L
M	UH	EH	VH	H	MH	M	ML	L	VL
ML	EH	VH	H	MH	M	ML	L	VL	EL
L	VH	H	MH	M	ML	L	VL	EL	UL
VL	H	MH	M	ML	L	VL	EL	UL	AL

Table 2 Desired Central Values of the Fuzzy System Output

x_2^d \ x_1^d	UL	EL	VL	L	ML	M	MH	H	VH
VH	6	6	6	5	4	3	2	1	0
H	6	6	5	4	3	2	1	0	-1
MH	6	5	4	3	2	1	0	-1	-2
M	5	4	3	2	1	0	-1	-2	-3
ML	4	3	2	1	0	-1	-2	-3	-4
L	3	2	1	0	-1	-2	-3	-4	-5
VL	2	1	0	-1	-2	-3	-4	-5	-6

In MSFs of input1, the sets in the middle ranges such as ML and M cover a wider area than the others since in the middle ranges the buffer occupancy is far from critical conditions and so the QP is kept unchanged or change slowly. On the other hand, where the buffer status is critical, the sets close to zero or one cover narrower ranges to allow faster changes of QP. In other words, while the normalized buffer occupancy is about 0.6 and the normalized consumed bits is close 1 so the inputs are close to the ideal condition and there is no need to change the QP. As the result, the desired central value corresponding to the such area is set equal to 0 and whatever we select the MSFs in such regions wider, so the QP will be kept unchanged in wider region. However, when the inputs are far from ideal condition and close to the critical regions such as normalized buffer occupancy close to 1 and 0, the QP should be changed abruptly to prevent buffer overflow and underflow. Since the abrupt QP change, increases the unwelcomed quality fluctuation, the MSFs should be narrow in order to limit abrupt QP fluctuation to the critical regions.

Finally, by using a singleton fuzzifier, a product interface engine, and a center average defuzzifier the output of the fuzzy system is computed by (8):

$$f^d(x_1^d, x_2^d) = \frac{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2} \bar{y}^{i_1 i_2}\, \mu_{A_1^{i_1}}(x_1^d).\mu_{A_2^{i_2}}(x_2^d)}{\sum_{i_1=1}^{N_1}\sum_{i_2=1}^{N_2} \mu_{A_1^{i_1}}(x_1^d).\mu_{A_2^{i_2}}(x_2^d)}, \quad (8)$$

where f^d denotes the output of the fuzzy system for the d^{th} layer. $\{A_i^1, A_i^2, A_i^3, ...A_i^{N_1}\}_{i=1,2}$ stands for the input fuzzy set and $\bar{y}^{i_1 i_2}$ stands for the central desired value. N_1 and N_2 are the number of fuzzy sets for input x_1^d and x_2^d respectively. It is emphatic that all the aspects of the fuzzy controller are considered based on the expert experiences and experiments execution without performing an optimization process. However, the experimental results confirm that the performance of our proposed RCA is better than the others. Readers are referenced to [30] for more detailed information about fuzzy design and derivations. The output of the fuzzy system will be passed through a content-adaptive gain (G_f^d) which can be tuned in the range of (0.5 ~ 1) in order to adjust the control intensity accordingly to the video content as:

$$\Delta QPF_b^d = G_f^d \times f^d\left(x_1^d, x_2^d\right), \quad (9)$$

A higher gain is suitable for a video sequence with a lot of heterogeneous scenes and vice versa. The output of the fuzzy controller is used for computing the base QP as presented in equation (1).

2-3- SSIM-Based Quality Controller

The distortion model used in the HEVC framework is based on the error-sensitive metrics such as PSNR and MSE that are full-reference (FR) objective metrics. The main popularity reason for them is the calculation simplicity. They do not take the human visual system (HVS) characteristics into account while the output video quality is ultimately judged by human eyes. Researchers show that HVS is very sensitive to the structural information in the field of viewing. Therefore, it is better to use structural similarity-based metrics which have more correlation with HVS and estimate the output video quality more efficiently. SSIM is a structural similarity-based metric which attempts to extract structural information to evaluate the video quality.

To define SSIM, let α and β be the original signal and distorted signal respectively. μ_α and μ_β denote the mean of α and β, respectively which estimate the luminance. σ_α and σ_β stand for the variance that estimate the contrast, and $\sigma_{\alpha\beta}$ is the covariance of α and β which measures the non-linear similarity of α and β. By utilizing these parameters, the luminance (l), the contrast (c) and structure (s) comparison measures are defined as follow:

$$l(\alpha,\beta) = \frac{2\mu_\alpha\mu_\beta}{\mu_\alpha^2+\mu_\beta^2},$$

$$c(\alpha,\beta) = \frac{2\sigma_\alpha\sigma_\beta}{\sigma_\alpha^2+\sigma_\beta^2},$$

$$s(\alpha,\beta) = \frac{\sigma_{\alpha\beta}}{\sigma_\alpha\sigma_\beta}, \quad (10)$$

Then, the structural similarity measure is yielded (11) by combining these measures:

$$S(\alpha,\beta) = l(\alpha,\beta)c(\alpha,\beta)s(\alpha,\beta) = \frac{4\mu_\alpha\mu_\beta\sigma_{\alpha\beta}}{\left(\mu_\alpha^2+\mu_\beta^2\right)\left(\sigma_\alpha^2+\sigma_\beta^2\right)}, \quad (11)$$

The equation (11) is unstable and so it is modified to a new measure named SSIM as:

$$SSIM(\alpha,\beta) = \frac{\left(2\mu_\alpha\mu_\beta+C_1\right)\left(2\sigma_{\alpha\beta}+C_2\right)}{\left(\mu_\alpha^2+\mu_\beta^2+C_1\right)\left(\sigma_\alpha^2+\sigma_\beta^2+C_2\right)}, \quad (12)$$

where C_1 and C_2 are given by (13):

$$C_1 = (k_1 L)^2 \ , \ C_2 = (k_2 L)^2 , \tag{13}$$

where L is the dynamic range of pixel values set to 255 for 8-bit videos. k_1 and k_2 are two constants set to 0.01 and 0.03, respectively. Readers are referenced to [24, 31, 32] for more details.

In this paper, we take the advantages of SSIM as a quality metric and propose an SSIM-based quality controller to improve the performance of our algorithm. Our quality controller uses the SSIM of the compressed video in each layer as a feedback signal and calculates a QP change (ΔQPQ_b^d) in the range of (-2 ~ 2) for each layer as output. The relation between the input and output of our quality controller is represented in (14):

$$\Delta QPQ_b^d = G_q^d \times \overline{QP^d} \left(\overline{SSIM^d} - SSIM_{b-1}^d \right), \tag{14}$$

where the $\overline{QP^d}$ and $\overline{SSIM^d}$ denote the average QP and the average SSIM, respectively for all previously encoded frames at the same layer. $SSIM_{b-1}^d$ stand for the average SSIM of previously encoded GOP. G_q^d is a constant gain which can be used to adjust the control intensity. The output of the quality controller is used for computing the base QP as presented in equation (1).

3- Experimental Results

To evaluate the performance of proposed RCA, we implemented our proposed algorithm on the SHVC standard reference software SHM-12.1 [33] and executed a set of experiments. The random access scalable configuration with three layers including a base layer, a spatial 2.x layer, and an SNR layer is used for the experiments. The reason for using a spatial and an SNR layer as enhancement layers is to show the performance of our algorithm on both types of scalable layers. Each enhancement layer uses only the previous layer in interlayer processing. In order to configure the RCA, the size of each buffer is chosen equal to 1.5 seconds buffering of a bitstream with the aggregated target bit rate. Moreover, the gains of the fuzzy rate controller and the quality controller were set to 0.65 and 0.7, respectively.

We used a set of well-known sequences such as Keiba, RaceHorses, BQMall, BasketballDrill, PartyScene, Kimono, and ParkScene in our experiments. The proposed RCA is targeted for long-term rate-controlling, so we concatenated short test sequences to make longer sequences, suitable for our experiments. KR, BP, and KP are the abbreviations for the name of concatenated sequences Keiba to RaceHorses, BasketballDrill to PartyScene, and Kimono to ParkScene, respectively.

In video encoding with a constant QP (CQP), there is no control over the bit rate and therefore, there is no guarantee for the buffer constraint to be obeyed especially for a long time. However, CQP encoding provides smooth and higher visual quality for compressed video. On the other hand, the λ-domain RCA implemented in the reference software is supposed to produce a constant bit rate suitable for low-delay applications. From the operating region point of view, a high-delay RCA should operate in a region between low-delay algorithms and CQP case. Hence, we selected the λ-domain RCA and CQP cases in order to compare our proposed algorithm with them from the rate control and video quality points of view. These two algorithms are compared with the proposed algorithm in terms of mean QP, mean PSNR and mean SSIM. PSNR is a signal fidelity metric that measures the correlation between the original video and the encoded one. The higher PSNR means higher quality. As discussed in the previous section, SSIM measures the structural similarity between the original video and the processed one and reports the similarity value in the range of (0~1). The closer SSIM to 1 means higher quality. Moreover, we introduced a metric namely Mean Absolute Gradient (MAG) as a fluctuation metric to compare the algorithms in terms of fluctuations on QP, PSNR, and SSIM. The MAG on the variable θ^d is defined as:

$$MAG(\theta^d) = \frac{1}{M-1} \sum_{a=0}^{M-1} \left| \theta_{a+1}^d - \theta_a^d \right|, \tag{15}$$

where the variable θ^d can be substituted by QP, PSNR or SSIM in order to compute the MAG of these metrics. M denotes the number of encoded frames and stands for the index of each frame in display order. The MAG of PSNR and SSIM measure that how much the quality and structural similarity of consequent frames are correlated. Small MAG of PSNR and SSIM means that the quality of the output video has smooth fluctuation and a more pleasant video display is provided for the user.

To evaluate the performance of our RCA from the buffering constraints point of view, the encoded sequences are compared in term of minimum initial buffering delay which is formulated in (16):

$$Delay^d = \frac{0.6 \times \left(BO_{\max}^d - BO_{\min}^d \right)}{\sum_{j=0}^{d} TR^j}, \tag{16}$$

where a higher delay means more variations in bitrate. A high delay value can be interpreted as overflow or underflow or both of them. However, low delay values cannot be necessarily interpreted as perfect control. In other words, (16) measures the time that should be passed until 60% of the buffer space be filled with the incoming bits. For perfect control, the following constraints must be obeyed for all GOPs in all layers:

$$BO_{\max}^d \le BS^d \quad \text{AND} \quad BO_{\min}^d \ge 0 , \tag{17}$$

From the rate control point of view, it is notable that in all experiments, the proposed RCA completely obeyed the buffering constraints with neither buffer overflow nor underflow and successfully achieved the target rate. On the other hand, the virtual buffer simulated for CQP and λ-domain RCAs has overflow or underflow in several cases.
The test sequences were encoded by the algorithms in four operation points according to the SHM common test conditions [34] and for the lack of space, only a part of numerical experimental results are represented in Table 3.
According to Table 3 by averaging PSNR mean over all layers of the test sequences, the values of 37.38, 37.46, and 37.29 are resulted by the CQP, the proposed (S-F), and the λ-domain (LAD) RCAs, respectively.

Table 3(a) Comparison Simulation Results of S-F with CQP and LAD

Sequences	Layer ID	RCA	PSNR (dB)		SSIM	
			Mean	MAG	Mean	MAG
KR	BL	CQP	34.74	1.47	0.932	0.013
		S-F	34.91	1.44	0.926	0.013
		LAD	34.75	1.81	0.921	0.016
	EL1	CQP	35.32	1.32	0.928	0.011
		S-F	35.44	1.24	0.922	0.011
		LAD	35.35	1.81	0.918	0.016
	EL2	CQP	37.33	1.68	0.948	0.011
		S-F	37.42	1.55	0.945	0.011
		LAD	37.27	2.20	0.941	0.015
BQMall	BL	CQP	32.64	0.58	0.928	0.006
		S-F	32.94	0.60	0.929	0.006
		LAD	32.80	0.64	0.925	0.007
	EL1	CQP	33.65	0.47	0.915	0.005
		S-F	33.82	0.48	0.914	0.005
		LAD	33.50	0.58	0.907	0.007
	EL2	CQP	35.75	0.61	0.94	0.005
		S-F	35.87	0.61	0.939	0.005
		LAD	35.59	0.69	0.935	0.006
BP	BL	CQP	39.75	1.44	0.980	0.004
		S-F	39.84	1.49	0.980	0.005
		LAD	39.67	1.87	0.979	0.006
	EL1	CQP	39.38	1.20	0.970	0.004
		S-F	39.52	1.14	0.970	0.004
		LAD	39.23	1.75	0.967	0.007
	EL2	CQP	41.80	1.70	0.981	0.004
		S-F	41.84	1.75	0.981	0.005
		LAD	41.62	2.48	0.979	0.007
KP	BL	CQP	38.45	1.00	0.958	0.007
		S-F	38.46	0.97	0.957	0.007
		LAD	38.39	0.85	0.956	0.006
	EL1	CQP	39.04	0.73	0.946	0.006
		S-F	38.91	0.73	0.944	0.006
		LAD	38.86	0.73	0.943	0.006
	EL2	CQP	40.70	0.84	0.958	0.006
		S-F	40.57	0.82	0.957	0.006
		LAD	40.47	0.98	0.957	0.007
Total-Average		CQP	37.38	1.09	0.949	0.007
		S-F	37.46	1.07	0.947	0.007
		LAD	37.29	1.37	0.944	0.009

The results show that our algorithm provides a higher video quality level than the CQP and the λ-domain algorithms in terms of PSNR. Also, by averaging the MAG of PSNR over the tested sequences the values of 1.09, 1.07, and 1.37 are resulted by the CQP, the S-F, and the λ-domain RCAs, respectively. According to these results, our algorithm has provided less fluctuation in PSNR than the anchors and so it provides more constant visual quality. Moreover, by averaging the SSIM values over the test sequences, the values of 0.949, 0.947, and 0.944 are resulted by the CQP, the S-F, and the λ-domain RCAs, respectively. That means the performance of the

Table 3(b) Comparison Simulation Results of S-F with CQP and LAD

Sequences	Layer ID	RCA	QP		Delay (Sec)	Average Bit-Rate (kbps)
			Mean	MAG		
KR	BL	CQP	33.10	1.83	1.61	297.81
		S-F	32.86	1.83	0.46	294.89
		LAD	36.05	7.27	0.53	298.02
	EL1	CQP	33.10	1.83	2.12	876.02
		S-F	32.79	1.82	0.59	858.18
		LAD	35.80	7.38	0.65	876.44
	EL2	CQP	29.10	1.82	2.24	972.15
		S-F	28.76	1.81	0.47	962.25
		LAD	31.43	7.31	0.63	973.07
BQMall	BL	CQP	37.10	1.84	0.82	252.94
		S-F	36.91	1.85	0.37	257.38
		LAD	39.77	5.11	0.24	253.66
	EL1	CQP	37.10	1.84	0.91	570.83
		S-F	36.87	1.85	0.44	577.78
		LAD	39.19	5.52	0.27	571.75
	EL2	CQP	33.10	1.83	0.88	704.08
		S-F	32.77	1.84	0.35	705.15
		LAD	35.35	5.88	0.25	704.71
BP	BL	CQP	25.10	1.80	1.30	1461.06
		S-F	24.91	1.80	0.39	1472.64
		LAD	26.79	6.18	0.36	1461.01
	EL1	CQP	25.10	1.80	2.02	4804.95
		S-F	24.62	1.80	0.32	4810.19
		LAD	26.68	6.51	0.63	4809.08
	EL2	CQP	21.10	1.80	2.01	5358.12
		S-F	20.72	1.80	0.34	5288.33
		LAD	22.63	6.32	0.66	5359.77
KP	BL	CQP	29.10	1.83	1.22	1089.25
		S-F	29.08	1.81	0.54	1073.22
		LAD	30.54	3.95	0.79	1089.88
	EL1	CQP	29.10	1.83	1.20	2779.86
		S-F	29.19	1.81	0.60	2723.22
		LAD	30.60	4.31	0.83	2782.28
	EL2	CQP	25.10	1.82	1.36	3599.17
		S-F	25.10	1.8	0.54	3585.63
		LAD	26.75	4.90	0.87	3601.02
Total-Average		CQP	29.77	1.82	1.47	1897.19
		S-F	29.55	1.82	0.45	1884.07
		LAD	31.80	5.89	0.56	1898.39

proposed algorithm in terms of the visual quality of the compressed video is between those of the CQP and the λ-domain RCA as expected. The CQP algorithm uses a constant QP over the whole sequence while our proposed algorithm needs to vary the QP in order to control the bit rate and buffer state so the CQP outperforms the proposed algorithm. However, smooth control of QP by the proposed algorithm provides high visual quality for compressed video. Furthermore, by averaging the MAG of SSIM the values of 0.007, 0,007, and 0.009 are resulted by the CQP, the S-F, and the λ-domain RCAs, respectively. That means a similar performance in terms of visual quality smoothness for the proposed and the CQP algorithms. According to the average values for QP mean (29.77, 29.55, 31.80) and QP MAG (1.82, 1.82, 5.89) in the table, the proposed RCA provided a QP mean lower than those of CQP and λ-domain RCAs while in term of QP MAG the proposed RCA performs similar to the CQP and much better than the λ-domain RCA. From the initial buffering delay point of view, the average values of 1.47, 0.45, and 0.56 are resulted by the CQP, the S-F, and the λ-domain RCAs, respectively. According to these results, the proposed algorithm provided a lower initial buffering delay than that of the CQP case and even lower than that of the λ-domain RCA as a constant bit rate algorithm. The point is that the λ-domain RCA failed to obey the buffer constraints in several cases in the experiments while and the buffering constraints are completely obeyed by the proposed RCA. For more investigations, we compared the proposed RCA with the CQP and the λ-domain RCAs in terms of the rate-distortion performance utilizing the Bjϕntegaard metrics. The test sequences were encoded by the algorithms in four operation points according to the SHM common test conditions [34] and the Bjϕntegaard Delta PSNR (BDPSNR) and Bjϕntegaard Delta Bit Rate (BDBR) are computed between the proposed algorithm and the anchor algorithms. The BDPSNR measures the average PSNR between two rate-distortion curves. Positive BDPSNR means quality enhancement and negative BDPSNR means quality degradation. The BDBR measures the average bit rate between two rate-distortion curves. Negative BDBR means bit rate saving and positive BDBR is interpreted as bit rate wasting. The comparison results are presented in Table 4.

Also, the algorithms were compared in terms of the Bjϕntegaard Delta SSIM (BDSSIM) and Bjϕntegaard Delta Bit Rate (BDBR) and provided results are presented in Table 5. According to the average results reported in Table. 4, our RCA performs better than the anchor algorithms in terms of Rate-PSNR. Moreover, the reported average results in Table. 5 show that our RCA performs close to the CQP case (-0.0011, 3.8004) and better than the λ-domain RCA (0.0023, -6.0229) in terms of Rate-SSIM. As sample graphical results, Fig. 3-5 show the buffer occupancy (BO), PSNR, SSIM and QP graphs on GOP-based for the three layers of the KP test sequence. As presented in the figures, the buffer occupancy (BO) graphs resulted by the CQP and λ-domain RCAs show several buffer overflows and underflows while for the proposed RCA the buffer has neither overflow nor underflow. As shown in the graphs, the proposed algorithm efficiently uses the buffer size in order to reduce the QP and perceptual quality fluctuations. Also, strong correlations between the graphs of proposed RCA and corresponding graphs of CQP can be seen in the figures that mean a high performance for the proposed RCA close to the CQP encoding.

Table 4 Rate-Distortion Performance Comparison between S-F, CQP, and λ-Domain in Term of PSNR

Sequence Name	Layer ID	PSNR			
		S-F Vs. CQP		S-F Vs. LAD	
		BDPSNR	BDBR	BDPSNR	BDBR
KR	BL	0.188	-3.782	0.215	-4.113
	EL1	0.087	-2.631	0.088	-2.693
	EL2	0.058	-1.695	0.145	-4.071
BQMall	BL	0.168	-3.048	0.087	-1.612
	EL1	0.019	-0.583	0.164	-4.415
	EL2	-0.061	1.995	0.117	-3.590
BP	BL	0.126	-2.486	0.061	-1.195
	EL1	0.212	-5.132	0.244	-5.873
	EL2	0.158	-3.444	0.194	-4.227
KP	BL	0.077	-1.833	0.145	-3.457
	EL1	-0.046	1.552	0.133	-4.518
	EL2	-0.105	4.387	0.092	-3.299
Total-Average		0.073	-1.392	0.140	-3.589

Table. 5 Rate-Distortion Performance Comparison between S-F, CQP, and λ-Domain in Term of SSIM

Sequence Name	Layer ID	SSIM			
		S-F Vs. CQP		S-F Vs. LAD	
		BDSSIM	BDBR	BDSSIM	BDBR
KR	BL	-0.0044	8.0664	0.0042	-7.2240
	EL1	-0.0038	11.1120	0.0027	-7.2125
	EL2	-0.0026	10.8361	0.0025	-8.8341
BQMall	BL	-0.0010	2.6213	0.0013	-3.5838
	EL1	-0.0011	3.6879	0.0023	-6.7775
	EL2	-0.0010	5.2484	0.0009	-4.5998
BP	BL	0.0001	-0.2108	0.0012	-2.8484
	EL1	0.0022	-4.4932	0.0045	-8.8604
	EL2	0.0012	-3.2940	0.0024	-6.5988
KP	BL	-0.0008	1.8461	0.0023	-4.8108
	EL1	-0.0012	3.6716	0.0025	-7.0958
	EL2	-0.0013	6.5138	0.0010	-3.8290
Total-Average		-0.0011	3.8004	0.0023	-6.0229

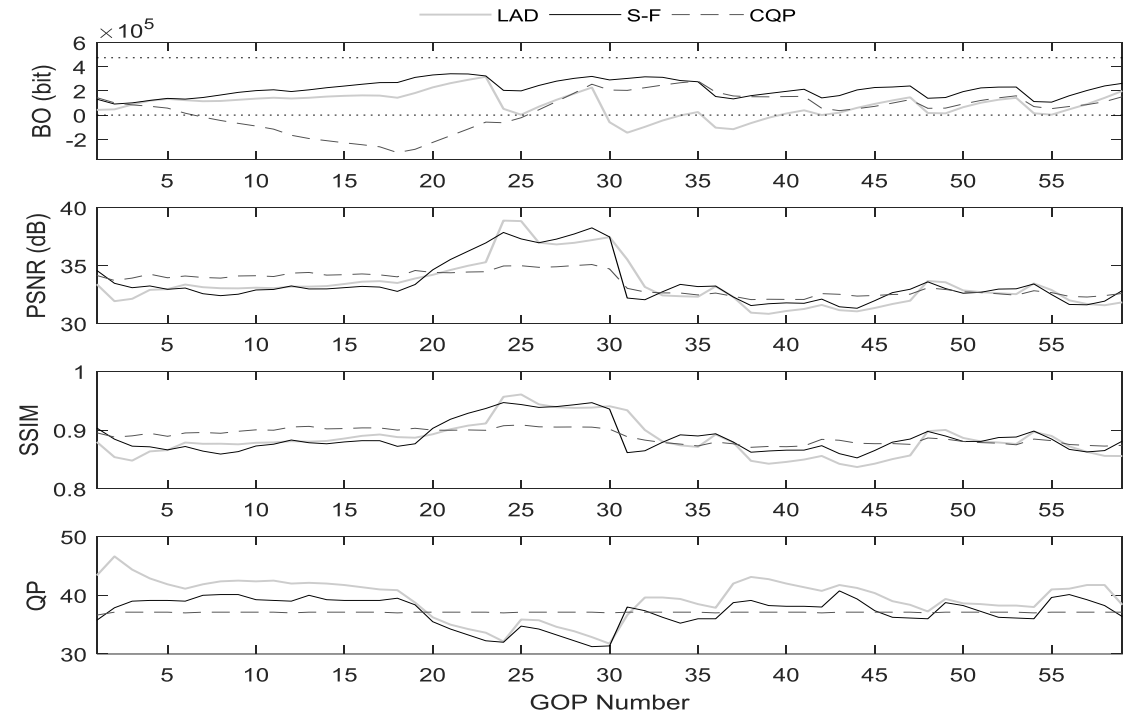

Fig. 3 Buffer Occupancy, PSNR, SSIM and QP Graphs of Base Layer

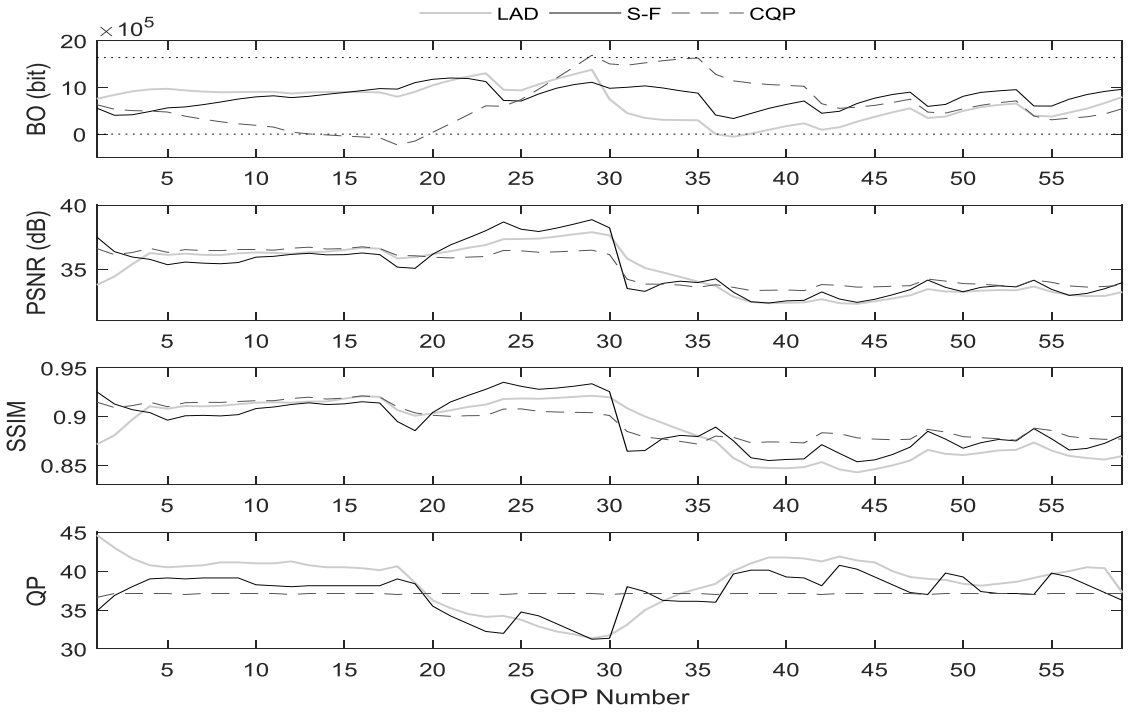

Fig. 4 Buffer Occupancy, PSNR, SSIM, and QP Graphs of Enhancement Layer1

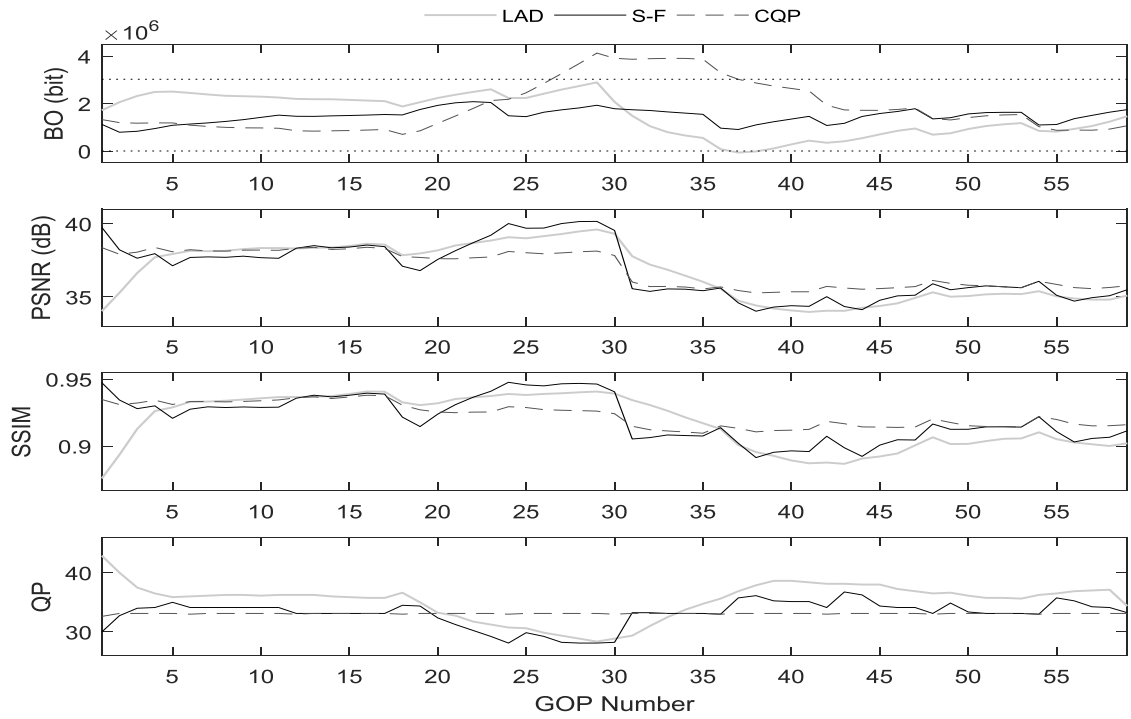

Fig. 5 Buffer Occupancy, PSNR, SSIM, and QP Graphs of Enhancement Layer2

4- Conclusions

In this paper, we proposed a video rate controller targeted for VBR applications of the scalable extension of HEVC standard, which is able to control the bit rate and buffer state in all types of scalable layers and it consists of a fuzzy rate controller and an SSIM-based quality controller. The fuzzy controller controls the bit rate while the attempt is to minimize the QP fluctuations and the quality controller smooths the SSIM quality metric over video frames.

Both controllers operate on the GOP level. The proposed rate control algorithm was implemented in the standard reference software and a comprehensive set of experiments was executed. According to the experimental results, the rate and buffering constraints are completely obeyed by the proposed algorithm. Also, from the rate-PSNR performance point of view, the proposed algorithm performs better than the CQP and λ-domain RCAs. Moreover, from the rate-SSIM performance point of view, the proposed algorithm performs better than the λ-domain RCA and close to the CQP case. Furthermore, from the QP and SSIM smoothness point of view the proposed RCA performs better than the λ-domain RCA and very close to the CQP encoding.

References

[1] V. Sze, M. Budagavi, and G.J. Sullivan, High Efficiency Video Coding (Hevc), Integrated Circuit and Systems, Algorithms and Architectures. Springer, 2014, pp. 1-375.

[2] G.J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (Hevc) Standard", IEEE Transactions on Circuits and Systems for Video Technology. vol. 22, no. 12. 2012, pp. 1649-1668.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H. 264/Avc Standard", IEEE Transactions on Circuits and Systems for Video Technology. vol. 17, no. 9. 2007, pp. 1103-1120.

[4] J.M. Boyce, Y. Ye, J. Chen, and A.K. Ramasubramonian, "Overview of Shvc: Scalable Extensions of the High Efficiency Video Coding Standard", IEEE Transactions on Circuits and Systems for Video Technology. vol. 26, no. 1. 2016, pp. 20-34.

[5] G.J. Sullivan, J.M. Boyce, Y. Chen, J.-R. Ohm, C.A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding (Hevc)", IEEE Journal of Selected Topics in Signal Processing. vol. 7, no. 6. 2013, pp. 1001-1016.

[6] M. Wien, High Efficiency Video Coding, Coding Tools and specification. 2015.

[7] H. Sun, T. Chiang, and X. Chen, Digital Video Transcoding for Transmission and Storage, CRC press, 2004.

[8] B. Li, H. Li, L. Li, and J. Zhang, "λ-Domain Rate Control Algorithm for High Efficiency Video Coding", IEEE

transactions on image processing. vol. 23, no. 9. 2014, pp. 3841-3854.
[9] MARZUKI, I., AHN, Y.-J. & SIM, D. 2017. Tile-level rate control for tile-parallelization HEVC encoders. Journal of Real-Time Image Processing, 1-19.
[10] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-Gop Based Rate Control for High Efficiency Video Coding", IEEE Journal of Selected Topics in Signal Processing. vol. 7, no. 6. 2013, pp. 1101-1111.
[11] H. Choi, J. Yoo, J. Nam, D. Sim, and I.V. Bajic, "Pixel-Wise Unified Rate-Quantization Model for Multi-Level Rate Control", IEEE Journal of Selected Topics in Signal Processing. vol. 7, no. 6. 2013, pp. 1112-1123.
[12] C.-W. Seo, J.-H. Moon, and J.-K. Han, "Rate Control for Consistent Objective Quality in High Efficiency Video Coding", IEEE transactions on image processing. vol. 22, no. 6. 2013, pp. 2442-2454.
[13] B. Lee, M. Kim, and T.Q. Nguyen, "A Frame-Level Rate Control Scheme Based on Texture and Nontexture Rate Models for High Efficiency Video Coding", IEEE Transactions on Circuits and Systems for Video Technology. vol. 24, no. 3. 2014, pp. 465-479.
[14] M. Wang, K.N. Ngan, and H. Li, "An Efficient Frame-Content Based Intra Frame Rate Control for High Efficiency Video Coding", IEEE Signal Processing Letters. vol. 22, no. 7. 2015, pp. 896-900.
[15] M. de-Frutos-López, J.L. González-de-Suso, S. Sanz-Rodríguez, C. Peláez-Moreno, and F. Díaz-de-María, "Two-Level Sliding-Window Vbr Control Algorithm for Video on Demand Streaming", Signal processing: Image communication. vol. 36. 2015, pp. 1-13.
[16] M. Rezaei, M.M. Hannuksela, and M. Gabbouj, "Semi-Fuzzy Rate Controller for Variable Bit Rate Video", IEEE Transactions on Circuits and Systems for Video Technology. vol. 18, no. 5. 2008, pp. 633-645.
[17] D. Fani and M. Rezaei, "A Gop-Level Fuzzy Rate Control Algorithm for High-Delay Applications of Hevc", Signal, Image and Video Processing. vol. 10, no. 7. 2016, pp. 1183-1191.
[18] R. Kamran, M. Rezaei, and D. Fani, "A Frame Level Fuzzy Video Rate Controller for Variable Bit Rate Applications of Hevc", Journal of Intelligent & Fuzzy Systems. vol. 30, no. 3. 2016, pp. 1367-1375.
[19] FANI, D. & REZAEI, M. 2017. Novel PID-Fuzzy Video Rate Controller for High-Delay Applications of the HEVC Standard. IEEE Transactions on Circuits and Systems for Video Technology, 28, 1379-1389.
[20] L. Li, B. Li, D. Liu, and H. Li, "λ-Domain Rate Control Algorithm for Hevc Scalable Extension", IEEE Transactions on Multimedia. vol. 18, no. 10. 2016, pp. 2023-2039.
[21] L. Li, B. Li, and H. Li, Rate Control by R-Lambda Model for Shvc. vol. JCTVC-M0037. 2013.
[22] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "Adaptive Rate Control Algorithm for Shvc: Application to Hd/Uhd", in Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on. 2016, IEEE. p. 1382-1386.
[23] F. Raufmehr and M. Rezaei, "Fuzzy Logic-Based Scalable Video Rate Control Algorithm for High-Delay Applications of Scalable High-Efficiency Video Coding", Journal of Electronic Imaging. vol. 27, no. 4. 2018, p. 043013.
[24] S. Akramullah, Digital Video Concepts, Methods, and Metrics: Quality, Compression, Performance, and Power Trade-Off Analysis, Apress, 2014.
[25] H. Zhao, W. Xie, Y. Zhang, L. Yu, and A. Men, "An Ssim-Motivated Lcu-Level Rate Control Algorithm for Hevc", in Picture Coding Symposium (PCS), 2013. 2013, IEEE. p. 85-88.
[26] H. Zeng, A. Yang, K.N. Ngan, and M. Wang, "Perceptual Sensitivity-Based Rate Control Method for High Efficiency Video Coding", Multimedia tools and applications. vol. 75, no. 17. 2016, pp. 10383-10396.
[27] GAO, W., KWONG, S., ZHOU, Y. & YUAN, H. 2016. SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation. IEEE Transactions on Multimedia, 18, 988-999.
[28] S. Wang, A. Rehman, K. Zeng, J. Wang, and Z. Wang, "Ssim-Motivated Two-Pass Vbr Coding for Hevc", IEEE Transactions on Circuits and Systems for Video Technology. vol. 27, no. 10. 2017, pp. 2189-2203.
[29] ZUPANCIC, I., NACCARI, M., MRAK, M. & IZQUIERDO, E. 2016. Two-pass rate control for improved quality of experience in UHDTV delivery. IEEE Journal of Selected Topics in Signal Processing, 11, 167-179.
[30] L.-X. Wang, A Course in Fuzzy Systems, Prentice-Hall press, USA, 1999.
[31] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE transactions on image processing. vol. 13, no. 4. 2004, pp. 600-612.
[32] Z. Wang, L. Lu, and A.C. Bovik, "Video Quality Assessment Based on Structural Distortion Measurement", Signal processing: Image communication. vol. 19, no. 2. 2004, pp. 121-132.
[33]Http://Hevc.Hhi.Fraunhofer.De/Svn/Svn_Shvcsoftware/Tags/Shm-11.0
[34] V. Seregin and Y. He, Common Shm Test Conditions and Software Reference Configurations. vol. JCTVC-Q1009. 2014.

Frahad Raufmehr received his BS degree in electronics engineering and his MSc degree in communications engineering from the University of Sistan and Baluchestan, Iran, in 2014 and 2016, respectively. He is currently pursuing his PhD in electronics engineering.

Mehdi Rezaei (IEEE SM) received the B.S. and M.Sc. degrees in electronics engineering from the Amir Kabir University of Technology (Polytechnic of Tehran) and Tarbiat Modares University, Tehran, Iran, in 1992, and 1996, respectively, and also Ph.D. degree in Signal Processing from the Tampere University of Technology (TUT), Finland, in 2008. His research interests include multimedia signal processing and communications. He has published several papers in these fields. During his Ph.D. program, he had a close collaboration with the Nokia Research Center and he filled several patents. He also received the Nokia Foundation Award in 2005 and 2006. Between 2015, he worked as Deputy Dean for Education and Postgraduate Studies, and as Deputy Dean for Research in Faculty of Electrical and Computer Engineering, and also Dean of Education at the University of Sistan and Baluchestan, Iran. Now he is Associate Professor and the head of Communications Engineering Department, University of Sistan and Baluchestan, Iran.