ORIGINAL PAPER

Region-of-interest-based rate control scheme for high-efficiency video coding

MARWA MEDDEB, MARCO CAGNAZZO AND BÉATRICE PESQUET-POPESCU

This paper presents a novel rate control scheme designed for the newest high efficiency video coding (HEVC) standard, and aimed at enhancing the quality of regions of interest (ROI) for a videoconferencing system. It is designed to consider the different regions at both frame level and coding tree unit (CTU) level. The proposed approach allocates a higher bit rate to the region of interest while keeping the global bit rate close to the assigned target value. The ROIs, typically faces in this application, are automatically detected and each CTU is classified in a region of interest map. This binary map is given as input to the rate control algorithm and the bit allocation is made accordingly. The algorithm is tested, first, using the initial version of the controller introduced in HEVC test model (HM.10), then, extended in HM.13. In this work, we first investigate the impact of differentiated bit allocation between the two regions using a fixed bit rate ratio in intra-coded frames (I-frames) and Bidirectionally predicted frames (B-frames). Then, unit quantization parameters (QPs) are computed independently for CTUs of different regions. The proposed approach has been compared to the reference controller implemented in HM and to a ROI-based rate control algorithm initially proposed for H.264 that we adopted to HEVC and implemented in HM.9. Experimental results show that our scheme has comparable performances with the ROI-based controller proposed for H.264. It achieves accurate target bit rates and provides an improvement in region of interest quality, both in objective metrics (up to 2 dB in PSNR) and based on subjective quality evaluation.

I. INTRODUCTION

Rate control (RC) is an important tool that helps to deal with bit rate and compressed media quality fluctuations. RC methods have been widely studied and suitable schemes have been developed for specific applications [1]. This problem is also related to challenging issues such as resource availability, computational complexity and real-time [2]. More precisely, we consider RC for a specific class of applications, namely video conferencing. In this context, one of the most interesting issues to focus on is the quality enhancement of regions of interest.

Indeed, in various fields such as video conferencing systems, video surveillance, and telemedicine, the subjective visual quality mainly depends on some important areas, called regions of interest (ROIs). Therefore, many contributions have introduced rate control algorithms aiming the improvement of the quality in the ROIs. For example, in [3] a rate control scheme based on adjustable quality of the ROI has been proposed. The RC algorithm used the quadratic model implemented in H.264/AVC to compute for each

region a quantization parameter (QP) referring to a quality level chosen by the user. The same quadratic model is used in [4] to compute the QP of each macroblock and then adjust it referring to an input saliency map and to the number of bits allocated to each region. For a video surveillance system, RC in [5] uses a linear rate-quantization (R-Q) model to decide the bit-stream length and then the QP of each region.

A different ROI-based method has been proposed in [6]. It uses a macroblock classification based on rate-distortion (R-D) characteristics to generate three kinds of regions (called basic units). A weighted bit allocation per region is performed with predetermined factors in heuristic ways. Finally, a linear rate-quantization stepsize (R-QS) model and a distortion-quantization stepsize (D-QS) model compute a QP per basic unit.

These techniques considered the R-Q linear model [5, 6] or the quadratic RC model [3, 4] and are useful for H.264/AVC implementations. Meanwhile, the new HEVC standard has been recently finalized by ITU-T and ISO/IEC [7] and many works have focused on rate control and developed new R-Q schemes for it. In the reference software, two different algorithms have been proposed. The first one is based on a quadratic model and the mean absolute difference (MAD) between the original and the reconstructed signal [8, 9]. In the second algorithm, an

Departement of TSI, Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, F-75634 Paris Cedex 13, France

Corresponding author:
Marco Cagnazzo
Email: cagnazzo@telecom-paristech.fr

R-λ model that takes into account the hierarchical coding structure has been adopted [10]. This model, initially introduced in HM.10, has been improved in a recent version of the reference software (HM.13). Adaptive bit allocation at the frame level has been introduced [11] by considering variable weights for each hierarchical level that depend on video content characteristics. Then, in [12], intra frame rate control has been modified by enabling basic unit level rate control.

Recent works on high-efficiency video coding (HEVC) have proposed bit allocation approaches that take into account coding units (CUs) texture. In [13], CUs are classified referring to their depth in the quad tree and their coding type. Texture-based rate models for HEVC have been developed according to signal characteristics in different CU depths and coding types. Rate models for three types of Cus of different texture levels have been constructed to deal with more complex content and to ensure more accurate rate control at the CU level.

All the above-mentioned RC algorithms, which have been developed for HEVC, do not take into account the importance of particular regions of the frame. Therefore, we propose a new rate control scheme for video conferencing systems which processes the faces (ROIs) and the background separately. We propose two versions for this algorithm. The first one is based on the model implemented in the reference software HM.10 [10] performing the R-λ model on B-frames only. While, the second one considers the modified RC model introduced in HM.13 [11] that performs RC for both I and B frames. The reference controller is enhanced with three main features; first, using an object detection method, we detect our ROI and generate automatically a binary map (ROI map). The target bit rate is allocated for each region considering a fixed weight. Then, the QP of each coding tree unit (CTU) is computed referring to the rate model of the corresponding region and the allocated bit budget. Finally, the proposed method considers independent R-D models for each region and different clipping values for QP variation, taking into account the importance of each part of the image. Overall, we show that the quality of the ROI is improved and the bit rate limit is respected. The proposed method has been compared to the ROI-based controller described in [4]. This algorithm based on a quadratic representation of the R-D model was implemented and tested for H.264/AVC. In our work, we adapt it to HEVC, implement it in HM.9 and compare its performance to our approach.

The paper is organized as follows. Section II introduces different methods for ROI detection and describes the general rate control problem. Then, Section III briefly reviews the main features of rate control for HEVC based on the R-λ model, compares it to the quadratic model and studies the evolution of the controller. The proposed algorithm is explained and detailed in Section IV. Moreover, the main modification of the adopted controller are detailed. In Section V, the experimental results related to both HM.10 and HM.13 integrations are presented and the controller

implemented in HM.9 is evaluated. Finally, conclusions and future research directions are given in Section VI.

II. RELATED WORKS

A) ROI detection algorithms

Many algorithms have been proposed for automatic ROI detection. They can be classified into two categories: bottom-up methods assume that human eyes skirt rapidly across the entire image and select small areas, while top-down methods suppose that people pay more attention to areas corresponding to semantic objects of the image [14]. Top-down approaches mainly consist in generating a saliency map taking into account the importance of semantic objects such as text, faces, eyes, etc.

One of the earliest works in face detection is a real-time system developed in [15] to emphasize the face region. The proposed method is based on a shape recognition algorithm. The system is able to detect and track human faces considering skin color segmentation and contour evaluation.

Viola and Jones object detector [16] is a famous and successful tool, widely used for face detection. For specific applications, such as video conference or supervision systems, this algorithm is appropriate as it has shown strong power in detecting faces, while for other applications, some improvements have been introduced to Viola and Jones algorithm by introducing new feature images. This framework used a set of Haar-like features in which each characteristic was described by a template. OpenCV library has included different implementations of Viola and Jones object detector algorithm [17].

In our work, as we focus on video conferencing systems, we used OpenCV library for face detection and ROI map generation. We do not aim at making a perfect segmentation of the face at the pixel level. The generated mask is done at the CTU level. Thus, a binary mask is generated to register if each CTU of the frame belongs to the ROI or not.

B) RC theory

The objective of RC is to achieve a target bit rate as close as possible to a given constant while ensuring minimum quality distortion. Knowing that quantization consists in reducing the bit rate of the compressed video signal, the major role of RC algorithms is thus to find for each transform coefficient the appropriate QP under the constraint $R_s(\text{QP}) \leq R_{max}$. The fixed bit budget is $R_{\max}$ and $R_s(\text{QP})$ is the number of coding bits for the source sample s. If we note D_s the distortion measure between the original and the constructed samples, the problem can be formulated as:

$$\min_{\text{QP}} D_s(\text{QP}) \text{ subject to } R_s(\text{QP}) \leq R_{\max}. \quad (1)$$

In video coding, RC usually incorporates rate-distortion optimization (RDO). Knowing the QP given by the rate

control, RDO consists in minimizing the cost

$$J = D(\text{QP}) + \lambda_{MODE} R(\text{QP}) \qquad (2)$$

to achieve the optimized mode decision of each CU. Using a Lagrange multiplier λ_{MODE} in (2), the distortion $D(\text{QP})$ is associated with the number of bits $R(\text{QP})$ to evaluate all possible coding modes and select the one minimizing J [18, 19].

Consequently, these problems need explicit models that relate the average bit rate to the QP. Several works have therefore been done in perceptual quality, for estimating the distortion, and in rate modeling. Different rate models have been developed, some of them based on simple linear expressions, others on more complex mathematical representations. For example, in [20], the traditional linear model that was employed in TM5 for high bit rate video coding is studied for HEVC:

$$R(\text{QP}) = \frac{C}{\text{QP}}, \qquad (3)$$

where C is the model parameter. The quadratic model represented as

$$R(\text{QP}) = \frac{C_1 \times \text{MAD}}{\text{QP}} + \frac{C_2 \times \text{MAD}}{\text{QP}^2}, \qquad (4)$$

where C_1 and C_2 are the model parameters and the MAD is the MAD between the original frame and the reconstructed one [8]. This model has been adopted in VM8 for MPEG4 [21], H.264/AVC [22], and also for HEVC [23].

The accuracy of these models has been enhanced by introducing the so-called complexity of the source, using the per pixel gradient value in the R-Q model in [24]. The sum of absolute transformed differences (SAD) has been adopted in [25]. In a different way, the RC was improved by considering a representation in the ρ domain [26] as proposed in [27] and by taking into account additional parameters, like the frame rate [28].

The most recent R-D model in the HEVC reference software is the R-λ model expressed as follows:

$$\lambda = \alpha R^\beta, \qquad (5)$$

where α and β are the model parameters [10]. We note that this model defines a relationship between the rate in bits per pixel, R, and the Lagrange parameter λ which is used in RDO to decide the coding mode. Using this R-λ model, λ is generated first, and then the QP at the frame level is computed. In our work, this model has been adopted for the frame level bit allocation and modified for our video conferencing system.

C) ROI and rate control

With rapid demands for ROI in applications like video conferencing, video surveillance and telemedicine, ROI-based rate control has gained increasing attention from researchers [3–6]. All these works are based on the controller implemented in the H.264/AVC reference software.

In [3], an ROI quality adjustable rate control algorithm has been proposed. Bit allocation is initially done according to user's interest level and available budget. The proposed quadratic R–D model defined in (4) considers the bit rate constraint and possible quality levels to define a QP margin. A number of bits is then allocated for each region and the QP is refined. In this scheme, ROI is processed first, and then the non-ROI areas. A QP is assigned for each region.

In [4], the same quadratic model is used and again faces are considered as ROIs. However, new features are introduced in this proposition. First, human psychovisual clues are used to compute a saliency map for each frame, which is used for rate control. A quality factor is defined and the bit budget is allocated for ROI and non-ROI separately. Finally, the quadratic model is used to assign a QP for each region considering a clipping range for smooth visual quality along the temporal direction and across region boundaries.

In [5], an ROI-based rate control was designed for traffic surveillance systems. A fast ROI extraction method for the real-time video compression is used to generate the ROI map. A linear function has expressed the relation between the bit-stream length and the quantization step (3). This model helps to predict the frame level bit allocation and the region level QP determination. In this work, the model is used for each block. Thus, a QP is computed for each macroblock.

In [6], a complete ROI-based controller is proposed. The scheme includes five steps, starting with region devising using the R-D characteristics of each macroblock. Macroblocks with similar characteristics are classified in the same basic unit and an overall bit allocation is performed using two linear models (R-QS and D-QS). A QP is computed for each basic unit. Finally, RDO is performed for each macroblock and models' parameters are updated as done in previous propositions.

The above-mentioned algorithms provide a bit rate repartition that takes into account the high priority of the ROI. They have been developed considering linear and quadratic models and implemented in the H.264/AVC JM software. In this paper, we propose a new ROI-based rate control scheme for HEVC characterized by several features. First, face detection is performed using Viola and Jones algorithm [16], an ROI map is generated and an ROI and non-ROI bit partition is determined accordingly. Second, the proposed model is used separately for ROI and non-ROI. Finally, a QP is assigned for each unit and clipped to keep quality smoothness.

The RC algorithm proposed in [4] is the most appropriate for HEVC. It is possible to adapt it to the HEVC controller, as it uses a quadratic model for QP computing which is not the case in reference [5] and keeps processing blocks in encoding order, which is not the case in the reference algorithm [3]. Consequently, the proposed ROI-based controller in [4] has been implemented in HM.9 and compared to our algorithm. A detailed description of these algorithms is given in the next section.
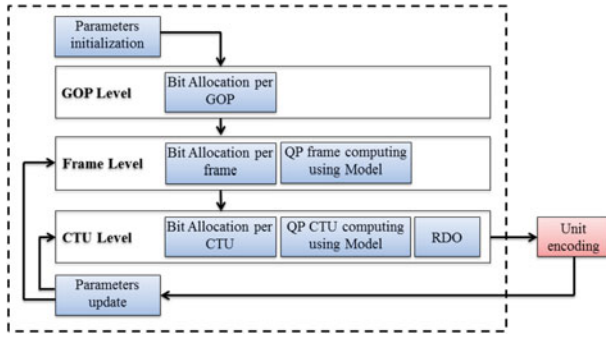
Fig. 1. RC scheme for HEVC.

III. RATE CONTROL FOR HEVC

In video coding, controllers have been designed to achieve the main goals of high coding efficiency and accurate matching of the target rate. Our approach aims at providing an ROI-based bit allocation between regions and achieves the controller goals (quality and budget). Thus, it is important to evaluate the key elements of the controller before introducing the ROI constraint. This section briefly describes rate control scheme for HEVC and different options at the frame and CTU levels.

A) R-λ-based rate control

1. R-λ SCHEME
As stated before, each model targets a specific video coding system under particular conditions. However, all the rate control methods aim at allocating the appropriate number of bits and at determining the QP of each CTU. The complete R-λ rate control scheme in HEVC can be represented as illustrated in Fig. 1:

As it can be seen from this figure, the controller operates at three main levels: group of pictures (GOPs), frame, and CTU [7]:

  (i) GOP level: the input parameters are the global target bit rate, the sequence frame rate, the GOP size and the virtual buffer occupancy. The rate control algorithm computes here an average number of bits per GOP.
  (ii) Frame level: considering the average number of allocated bits per frame, a target bit rate is fixed for the current frame. For B-frames, the bit allocation can introduce equal, hierarchical or adaptive weights, whereas for I-frames the initial budget is refined using a predefined multiplication weight. Then, the R-λ model is used to compute the frame QP.
  (iii) CTU level: the process is divided into three main parts. First, the required number of allocated bits for the CTU is computed using the frame budget, the cost of the coded CTUs of the frame and the complexity of the CTUs. The complexity is measured using the MAD [10] or the sum of absolute transformed differences (SATD) [12]. Second, the budget is used in the R-λ model to compute λ and then the QP of each CTU. The QP variation is clipped in a pre-defined

range. Finally, the last step is the RDO in order to find the optimized mode decision [29], referring to the obtained QP. The unit is then repartitioned, coded and all the parameters are updated.

2. COMPARISON BETWEEN THE R-λ MODEL AND QUADRATIC MODEL IN HEVC

Both quadratic and R-λ models have been used for rate control in HEVC. The first proposed controller is based on the unified quadratic model (URQ) described in (4) and has been introduced in HM.5, and then improved in later versions. It helps to reduce bit fluctuation and ensures a good quality encoding [8]. The R-λ model as represented in (5) has been introduced in HM.11 [10] and improved in HM.13 [11].

Comparative tests made to choose the appropriate model for our work show that global R-D performances are improved using the R-λ model. Referring to Fig. 2 the gain goes from -22.6 to -79.6% for Class E sequences [30] and using a low delay configuration with an intra period equal to 60.

Fig. 3 shows some per-frame bit cost comparing the used R-λ model and the old URQ model. For example, for the test sequence "Johnny" at different bit rates, R-λ model gives a better bit distribution over GOPs and a smoother repartition of the bit budget at the frame level.

B) GOP-level bit allocation

At the GOP level, bit allocation is performed as described in [10], taking into account the target bit rate R_{max}, the frame rate f and the number of frames in a GOP N_g. The target number of bits in a GOP are determined by:

$$T_g = N_g \left(\frac{(R_{\max}/f)(N'_s + S_w) - T'_s}{S_w} \right), \qquad (6)$$

where the smoothing window S_w is equal to 40, N'_s is the number of pictures already coded and T'_s is the bit cost of these pictures.

C) Frame-level bit allocation

Both inter and intra picture bit allocation are supported in the HEVC rate control algorithm, but the process is different. All I-frames belongs to the same level. Thus, the same factor is used to refine their allocated budget, while the cost of inter pictures is determined according to different weights w_p for the different hierarchical levels.

At the frame level, an initial budget is allocated per frame, using T_g computed in (6) and the bit cost of already coded pictures in the current GOP T'_g,

$$T_i = \frac{T_g - T'_g}{\sum_{j>p} w_j} w_p. \qquad (7)$$

1. WEIGHTED BIT ALLOCATION FOR INTER PICTURES
As said before, at the frame level, three main ways of bit allocation are possible. Equal and hierarchical bit allocations
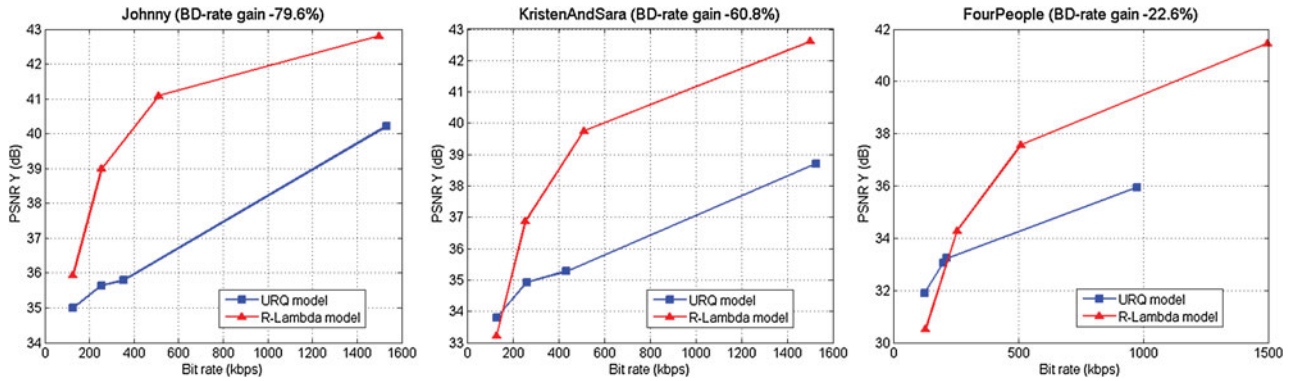
Fig. 2. R–D performances of the R-λ algorithm, compared URQ model.
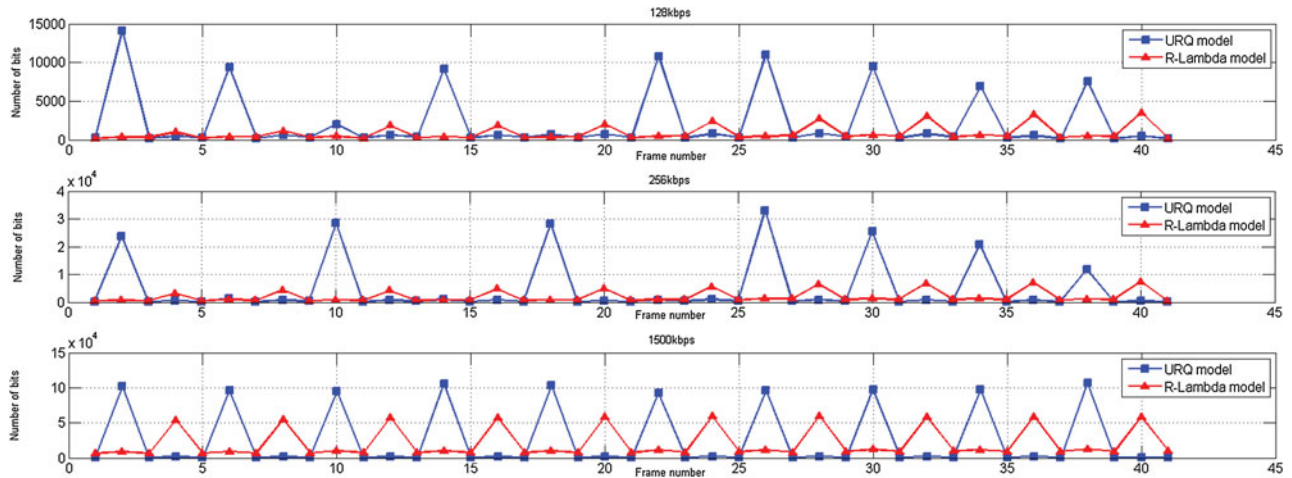
Fig. 3. Comparison of bit fluctuation per frame of R-λ and URQ models for sequence Johnny.

have been introduced in HM.10. Then, adaptive bit allocation has been added in the latest version of HEVC test model (HM.13) [11] to improve the model performance. Equal bit allocation method considers the same weight for all B-frames of the sequence. Hierarchical bit allocation consists in giving a predetermined weight to each frame B referring to its level in the GOP and the target bit rate. Using adaptive bit allocation importance weights are updated for each GOP considering the Lagrangian parameter λ computed as in (5) [11].

In Table 1, we compare the global performance of the controller using equal, hierarchical and adaptive bit allocation. We compute the R-D performance of the hierarchical method then the adaptive one compared to equal bit allocation. The comparison is made with low delay configuration and using test sequences of class E with video conference content [30]. Results show that hierarchical and adaptive methods are slightly better then the equal bit allocation and the adaptive allocation gives the best performance, with 1.6% of gain compared to the hierarchical one.

2. BUDGET REFINEMENT FOR INTRA PICTURES

In the R-λ controller implemented in HM.10, the refinement is done considering a weight W that depends on the number of bits per pixel as specified in Table 2 [10].

Table 1. R-D performance of R-λ algorithm using hierarchical and adaptive bit allocation, compared to equal bit allocation.

	Hierarchical bit allocation			Adaptive bit allocation		
	Y (%)	U (%)	V (%)	Y (%)	U (%)	V (%)
Class E	$-6,7$	$-12,1$	$-12,3$	$-8,3$	$-16,0$	$-16,0$
Enc Time		101			108	
Dec Time		99			110	

The final allocated budget per picture T_p is then:

$$T_p = W \times T_i, \tag{8}$$

where T_i is the initial allocated budget. In HM.13, intra picture bit allocation has been improved by replacing the old refinement method by:

$$T_p = a \times \left(\frac{C}{T_i}\right)^b \times T_i + 0.5, \tag{9}$$

where $a = 0.25$, $b = 0.5582$ and C is the complexity measure of the frame as defined in the next subsection [31].

D) CTU-level bit allocation

Large coding units (LCUs) or CTUs are the basic processing units used in HEVC standard to specify the decoding

Table 2. Intra bit allocation refinement weights.

Bit rate R	$R > 0.2$	$0.2 \geq R > 0.1$	$0.1 \geq R$
W	5	7	10

process. They are basically the replacement of macroblocks and blocks in prior standards. Each unit contains luma coding tree block (CTB) and the corresponding chroma CTBs and syntax elements. Bit allocation at the CTU level depends on three main features: CTU size, complexity measure, and R-λ model.

1. CTU SIZE
The size of a CTU can go up to 64×64 pixels in HEVC. At the unit level, rate control algorithm is applied to evaluate the average bit budget for each CTU and then compute a QP from the model. In our work, we consider CTUs of 64×64 pixels. First, larger sizes enable better encoding performances. Second, from our tests, the rate control algorithm shows a better matching of the target bitrate when the unit size is equal to 64×64.

2. COMPLEXITY MEASURE
To perform bit allocation, a weight w_B is computed for each CTU. In HM.10, the weight is estimated by the prediction error (in form of MAD) between the current unit p and the coded unit p' in the previous coded picture belonging to the same level [10]. The weight of each CU of index i is defined as:

$$w_B(i) = \left(\frac{1}{N} \sum_{j \in N} |p_j - p'_j| \right)^2, \tag{10}$$

where N is the number of pixels of the CTU.

In the latest version of the reference software (HM.13) [31], the CTUs weight for B-frames has been modified. It depends on the model parameters α_i and β_i at the CTU level, the λ of the picture and the number of pixels N. For a CTU of index i:

$$w_B(i) = N \left(\frac{\lambda}{\alpha_i} \right)^{\frac{1}{\beta_i}}. \tag{11}$$

Finally, the complexity measure for I-frame (CTU) is calculated by deriving the sum of absolute Hadamard transformed difference (SATD) as described in [12]:

$$\text{SATD} = \sum_{k=0}^{7} \sum_{\ell=0}^{7} |h_{k\ell}|, \tag{12}$$

where h_{kl} are the coefficients obtained after applying the Hadamard transform to the original 8×8 block. The weight $w_I(i)$ of a CU of index i is defined as the sum of SATD calculated for all 8×8 blocks within the CTU (N_b is the number of 8×8 units in the CTU).

$$w_I(i) = \sum_{j=0}^{N_b-1} \text{SATD}(j). \tag{13}$$

3. R-λ MODEL
The R-λ implementation introduced in HEVC reference software uses different methods of QP computing and bit allocation for the I and the B frames. In HM.10, bit allocation at the CTU level for intra frames was not considered. All the units have the same QP obtained at the frame level. For B-frames, the model introduced by (5) in Section II is used at the CTU level and its parameters α and β are updated after encoding each unit.

In HM.13, to better control the rate allocation of intra coded frames, the complexity measure defined in (12) and (13) is additionally taken into consideration in the R-λ model as follows:

$$\lambda_i = \alpha \left(\frac{w_I(i)}{R_i} \right)^{\beta}. \tag{14}$$

For a CTU of index i, λ_i depends on model parameters at the frame level. The parameters α and β remain constant for the entire frame; however, the number of allocated bits per pixel R_i is computed per CTU. Consequently, the model gives a λ and thus a QP for each CTU.

IV. PROPOSED APPROACH

The proposed approach is based on the R-λ model for HEVC. The relationship between R and λ represented by (5) in Section II is used to compute QP of the frame and each CTU of the image. This model has given better performance than the quadratic one [8, 9].

Our contribution proposes an ROI-based rate control algorithm where the CUs bit allocation depends on the number of bits allocated per region and on the weights of CTUs of the same region. The same process is done independently for units of different regions (ROI or non-ROI). In this section, we describe the initial proposed approach that has been implemented in HM.10 and how we adapted it to the latest version of HEVC test model 13 (HM.13). We focus on the two main steps of the rate control: the bit allocation at both the frame and CTU levels and the computation of QP by the proposed model for both I and B frames.

This section ends up with a detailed description of the adopted ROI-based controller [4] to HEVC URQ algorithm [9]. The implementation is done in HM.9 and results are illustrated in the next section.

A) ROI-based model scheme

Fig. 4 shows the proposed ROI-based rate control scheme. The first step consists in detecting the faces in the scene and generating automatically a binary ROI map per frame, which will be given as input to our controller. The target bit rates allocated for the GOP and the current frame are obtained using the reference algorithm described in [10] and improved in [12].

Then, the frame budget is divided into two parts according to a fixed factor K which is the *desired* ratio between
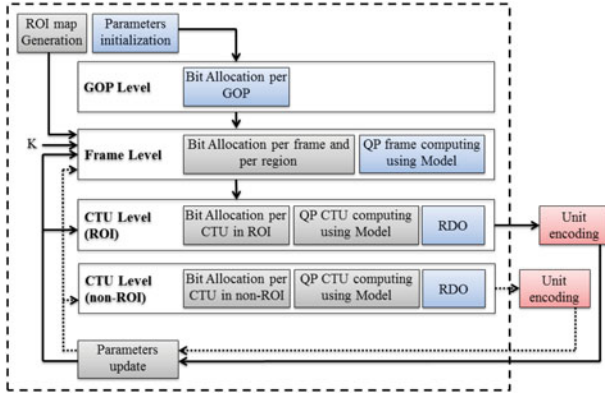
Fig. 4. ROI-based rate control scheme for HEVC.

the bit rate of the ROI and the bit rate of the rest of the frame (non-ROI). At the CTU level, the binary ROI map is used to make a separate bit allocation for CTUs of different regions. The R-λ model is then applied for each CTU using the allocated bit budget for the corresponding region (ROI or non-ROI). Once the CTU is encoded, the model parameters of the corresponding region are updated, and the next CTU is processed in a similar way.

In the first implementation of the controller (in HM.10), the described process is only used for B-frames of different hierarchical levels. Then, it was adapted to HM.13 and introduced in both I-frames and B-frames, considering some differences in complexity computing and model parameters update.

B) Region bit allocation for B-frames

We introduce the region bit allocation at two levels: at the frame level to initialize a target amount of bits for each region, and at the CTU level to make independent bit allocation of CTUs of different regions. At the frame level, the positive constant K is selected. It represents the desired ratio between the ROI and non-ROI bit rates:

$$R_r = K \times R_n, \tag{15}$$

where the subscript r denotes the ROI and n the non-ROI. We assume that the current number of allocated bits per frame T_p is the sum of the number of bits of the two regions, T_r for the ROI and T_n for the non-ROI:

$$T_p = T_r + T_n, \tag{16}$$

$$T_n = R_n \times M \times P_n, \tag{17}$$

where M is the total number of pixels of the frame and P_n the area of non-ROI. From (15), (16) and (17), the non-ROI bit rate R_n is computed as follows:

$$R_n = \frac{T_p}{M(1 + P_r(K - 1))}. \tag{18}$$

At the CTU level, the bit allocation for B-frames depends on the number of bits allocated per region and on the

weights of CTUs of the same region. For CTU of index i of the ROI, the allocated bits are:

$$T_r(i) = \frac{T_r - T_r'}{\sum_{j \in I_r} w_r(j)} w_r(i), \tag{19}$$

where T_r' is the effective number of bits of already encoded CTUs of the ROI, I_r is the set of indexes of ROI CTU that have not yet been coded, and $w_r(i)$ is the weight of the current CTU of the ROI computed referring to (10). The same process is applied independently to CTUs of the rest of the frame (non-ROI).

C) Region-independent rate control models

For B-frames, once the rate of each CTU is found, the QP is computed using the R-λ model. Our proposal separates the models of the different regions. Consequently, the model parameters of CTUs from the ROI r are independent from the ones of CTUs of the non-ROI n. In fact, we have two models; in ROI, using the effective number of bits per pixel $R_r(i)$ of each unit of index $i \in I_r$,

$$\lambda_r(i) = \alpha_r R_r(i)^{\beta_r} \tag{20}$$

and for CTUs from the non-ROI (of index $j \in I_n$), using the effective number of bits per pixel $R_n(i)$,

$$\lambda_n(j) = \alpha_n R_n(j)^{\beta_n}. \tag{21}$$

The model parameters are then updated separately. For the ROI, the parameters α_r and β_r are updated referring to the original rate control algorithm [10], as follows:

$$\lambda_r' = \alpha_r R_r'^{\beta_r}, \tag{22}$$

$$\alpha_r' = \alpha_r + 0.1(\ln \lambda_r - \ln \lambda_r')\alpha_r, \tag{23}$$

$$\beta_r' = \beta_r + 0.05(\ln \lambda_r - \ln \lambda_r') \ln R_r', \tag{24}$$

where α', β', and λ' are the updated values of α, β, and λ. In (22) and (24), R_r' is the effective number of bits per pixel after encoding the unit. The same update process is used for the CTUs of the non-ROI.

D) QP and λ variation

The last modification compared to the reference algorithm consists in considering new clipping ranges for λ and QP, at the CTU level. As we try to make independent QP computing for each region, the QP of the current CTU depends on the QP of the last CTU of the same region and the QP of the current frame. We allow a larger QP range than in the reference algorithm, to accommodate differences in quality between the ROI and the non-ROI. We define $\Delta QP_p > 2$ and $\Delta QP_u > 1$ that guarantees

$$QP_p - \Delta QP_p \leq QP_u \leq QP_p + \Delta QP_p, \tag{25}$$

$$QP_{u'} - \Delta QP_u \leq QP_u \leq QP_{u'} + \Delta QP_u, \tag{26}$$

where QP_u, QP_p, and $QP_{u'}$ are respectively the QPs of the current CTU, the current picture and the previously

encoded CTU of the same region. It is also possible to consider different clipping ranges for CTUs of different regions and use asymmetric clipping.

E) Extended version of ROI-based rate control algorithm

Modifications have been introduced to our initial approach taking into consideration the evolution of the controller in the new version of HEVC test model (HM.13). There are two main modifications in the new proposal: ROI bit allocation for frame B is adapted to the new version and ROI bit allocation for frame I at the CTU level is introduced.

1. B-FRAMES ROI BIT ALLOCATION
In the new version of the controller, the weight of a CTU is computed by (11). Thus, in our updated ROI-based controller the weight of a CTU from the ROI of index i is expressed as follows:

$$w_r(i) = N \left(\frac{\lambda_{\text{Pic}}}{\alpha_r} \right)^{\beta_r}, \qquad (27)$$

where α_r and β_r are the R-λ model parameters for CTUs of the ROI and λ_{Pic} is the current picture λ. This weight is then used to compute an initial target allocated bit rate $T_r(i)$:

$$T_r(i) = \frac{T_r \, w_r(i)}{\sum_{j \in I_r} w_r(j)}. \qquad (28)$$

The target allocated bits for a CTU $\widetilde{T}_r(i)$ takes into account $T_r(i)$, the allocated budget for the rest of CTUs of the same region, the effective number of bits of already encoded units of the ROI T'_r and a smoothing window W fixed at 4 in our simulations:

$$\widetilde{T}_r(i) = T_r(i) - \frac{\left(\sum_{\substack{j \in I_r \\ j \geq i}} T_r(j) - (T_r - T'_r) \right)}{W} + 0.5. \quad (29)$$

The number of bits per pixel for a CTU of the ROI is then:

$$R_r(i) = \frac{\widetilde{T}_r(i)}{N}. \qquad (30)$$

2. I-FRAMES ROI BIT ALLOCATION
At the frame level, the refinement of the initial number of bits is done referring to (9) then the K factor is considered to make ROI-based budget repartition as represented in (29) and compute T_r and T_n. At the CTU level, the weight of a unit is its cost and is calculated by deriving the SATD as described in Section III by (12) and (13). This weight is used to compute an initial target allocated bits $T_r(i)$ as in (28). Then, the number of bits left to encode the ith CTU $\widetilde{T}_r(i)$ is defined as:

$$\widetilde{T}_r(i) = (T_r - T'_r) + \frac{\left((T_r - T'_r) - \sum_{\substack{j \in I_r \\ j \geq i}} T_r(j) \right)(L_r - i)}{W}. \quad (31)$$

Finally, the number of bits per pixel for an intra CTU of the ROI is:

$$R_r(i) = \frac{\widetilde{T}_r(i) \, w_I(i)}{N \sum_{\substack{j \in I_r \\ j \geq i}} w_I(j)}. \qquad (32)$$

F) URQ ROI-based controller for HEVC

This section introduces a second ROI-based controller. As said before, the idea has been proposed in [4] for H.264 standard and it consists in estimating the bit count per region using a quadratic R-D model. We adopted this algorithm to the URQ controller introduced in [9]. The final version of the URQ ROI-based algorithm is then based on the model proposed in [4] and the controller implemented in HM.9 [32], but enhanced with several features.

1. BIT ALLOCATION PER REGION
At the frame level, separate bit allocation per region is performed. First, the initial budget fixed by the network is divided into two parts using a quality factor K as defined in (15) assigned by users or control systems. Target bit counts T_r and T_n are initialized to ROI and non-ROI referring to (16), then used for bit allocation at the frame and CTU levels.

The final target bit left budget $\widehat{T}_r(i)$ for CTU from the ROI is based on the remaining bits in ROI $(T_r - T'_r)$, the number of pixels in the current CTU $N(i)$ and the number of pixels left in ROI:

$$\widehat{T}_r(i) = \frac{(T_r - T'_r) \times N(i)}{\sum_{\substack{j \in I_r \\ j > i}} N(j)}. \qquad (33)$$

The final target bit occupancy $\widetilde{T}_r(i)$ for CTU from the ROI is computed using the initialized bit count in ROI and ROI virtual buffer occupancy $V_r(i)$:

$$\widetilde{T}_r(i) = T_r - \frac{V_r(i)}{U_r(i)}, \qquad (34)$$

where $U_r(i)$ is the number of units left in ROI after encoding CTU of index i.

The final bit budget is a weighted average of the target bit left and the target bit occupancy:

$$T_r(i) = \beta \times \widehat{T}_r(i) \times (1 - \beta) \times \widetilde{T}_r(i), \qquad (35)$$

where β is the weight defined in [9]. The same process is done for CTUs of the rest of the frame.

2. URQ MODEL FOR QP DETERMINATION
The strategy for intra pictures and non-reference frames is kept as described in the document [9] while the ROI-based URQ model is used at the CTU level for referenced B-frames. In this case, the final bit target $T_r^f(i)$ is refined as follows:

$$T_r^f(i) = T_r(i) \times \frac{w_B(i)}{\sum_{\substack{j \in I_r \\ j \geq i}} w_B(j)}, \qquad (36)$$

where $w_B(i)$ is the MAD of the current CTU as expressed in (10). After estimating this target bit count for the considered

CTU, the preliminary QP value is determined as in [9] by the quadratic model introduced in (4).

3. QP ADJUSTMENT

The QP obtained using the quadratic R-D model is then modified by considering the smoothness issues over the temporal and spatial domains. The four constraints proposed in [4] are used. All QPs are then clipped between 0 and 51 as proposed in URQ reference controller implemented in HM.9.

V. EXPERIMENTAL RESULTS

A) Test conditions

First, we implemented the proposed rate control scheme on HEVC test model 10 (HM.10) [33] and we evaluated the obtained results. Then we introduced the extended version on HM.13 [31, 34] by taking into account the evolution of the controller and compared the obtained results in the two cases. Finally, we implemented the URQ ROI-based model in HM.9 [32]. Performed tests help us evaluate and compare the performance of the proposed methods.

To compute a binary map as represented in Fig. 5, we used the same ROI detection method. We introduce HM Viola and Jones object detection algorithm [16].

Since video conferencing applications require low coding delay, all pictures were coded in display order. Three different configurations have been used to test the first and the second ROI-based controller: All-B, All-I, and an hybrid configuration that considers GOPs of B-frames and introduces an intra picture each second. In the first and the third algorithms (HM.10 and HM.9), I-frame bit allocation at the CTU level has not been yet introduced, so, all the frames were considered as B-frames except the first one (I-frame), while, for the extended version of our code in HM.13, we tested all the configurations.

Three HD 720p sequences from class E have been tested: "Johnny", "KristenAndSara", "FourPeople" [30]. As we can see in Fig. 5, the selected test sequences have typical video conferencing content and different characteristics, like number of faces and ROI size. We used different bit rates, budget partitioning per-region and QP ranges to evaluate the performance of our approach.

B) Implementation and performed tests

The introduced modifications have been done mainly in rate control class of the reference softwares HM.10 [33], HM.13 [31], and HM.9 [32]. A reference test "Ref" is performed using the rate control algorithm described in [10] and improved in [11]. While evaluating the URQ model the reference used is described in [9]. These first tests give us the reference performance: the ratio K between ROI bit rate and non-ROI bit rate, the bit budget used for encoding each region, the PSNR and the structural similarity (SSIM) index [35] of each region that goes from 0 to 100. Second, we activate all modified functions: we introduce a new bit
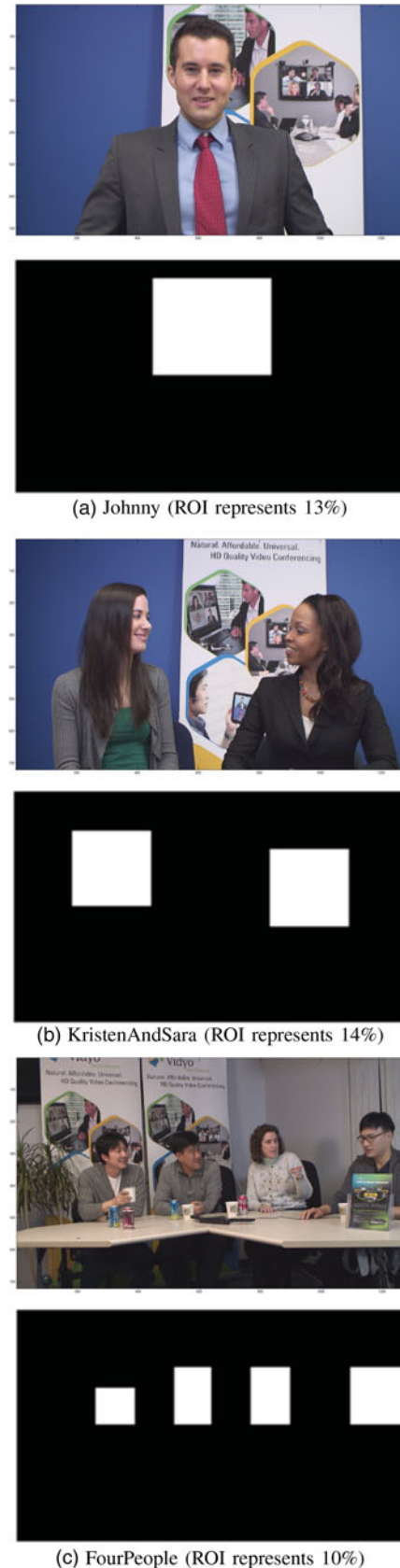

(a) Johnny (ROI represents 13%)

(b) KristenAndSara (ROI represents 14%)

(c) FourPeople (ROI represents 10%)

Fig. 5. Test sequences and ROI maps (a) Johnny (ROI represents 13%), (b) KristenAndSara (ROI represents 14%), (c) FourPeople (ROI represents 10%).

repartitioning between regions by fixing a factor K and a large QP margin. Then we perform an evaluation test of our method that we note "New".

C) Performance of ROI-based controller in HM.10

Table 3 summarizes the results of the performed test at 128 and 256 kbps. Both equal and hierarchical bit allocations are tested. The table shows that introducing a K factor for bit repartitioning between regions does not impair the R-D performance. We can increase the effective ratio comparing to the reference by keeping an output bit rate close to the assigned value. Moreover, the overall PSNR is practically the same as the reference encoder.

Now we examine the quality of ROI and non-ROI for different ratios K. In Table 3, ΔPSNR ROI is the difference in quality of the ROI using the proposed controller and the reference one and ΔSSIM ROI is the difference in similarity of the ROI using the proposed controller and the reference one (and the same for non-ROI). We notice that the overall quality of the ROIs is improved using different configurations but also different target rates. The global gain in the ROI goes from 0.5 to 0.7 dB in terms of PSNR and from 0.3 to 1.2 in terms of SSIM. However, as we reduce the number of allocated bits in the non-ROI, its quality decreases.

In Fig. 6, we plot ΔPSNR of the ROI and ΔPSNR of the non-ROI per GOP. Overall, the bigger is K the better is the global quality of the ROI in the sequence and the lower is the PSNR of the non-ROI. The quality of the ROI is improved in all the GOPs (and frames) while the quality of the non-ROI is slightly decreased. The curves show that for each region the difference in quality between the proposed scheme and the reference RC [10] is more important when K is bigger. This means that our method leads to allocate more bits to the ROI by improving its quality and respecting the bit rate constraint.

D) Performance of ROI-based controller in HM.13

1. INTRA PICTURE ROI-BASED ALGORITHM PERFORMANCE

Using the all intra configuration of the encoder, we tested the performance of the proposed algorithm. Three different rate points are used per sequence (640, 1280, and 2560 kbps). The budget constraint is respected and the global quality is not altered.

In intra case, units from the ROI are coded from other units of the non-ROI. Consequently, our novel bit repartition affects the non-ROI and so the ROI. We can see that in some cases the quality of the ROI decreases.

Table 3. Control accuracy comparison of the reference and the proposed controller for inter frames using HM.10.

Sequence	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		Δ PSNR (dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Equal bit allocation												
Johnny	128.01	127.89	36.48	36.04	92.76	92.07	5.82	10.41	0.76	−0.40	0.60	−0.90
	256.01	255.80	39.17	38.72	94.96	94.53	6.11	9.89	0.53	−0.46	0.30	−0.55
Kristen And Sara	128.04	128.02	33.96	33.74	92.20	91.94	3.35	5.10	0.70	−0.77	0.69	−0.43
	256.08	256.06	37.04	36.75	94.50	94.33	3.25	4.67	0.61	−0.68	0.43	−0.27
Four People	128.05	128.06	31.47	31.28	88.26	88.03	4.41	6.67	0.61	−0.33	1.22	−0.42
	256.07	256.06	34.48	34.27	92.28	92.09	4.33	6.16	0.61	−0.35	0.87	−0.33
Hierarchical bit allocation												
Johnny	128.96	127.73	37.15	36.64	93.46	92.74	5.47	9.27	0.65	−0.38	0.45	−0.91
	256.01	255.84	39.48	39.20	95.21	94.91	5.95	9.62	0.66	−0.34	0.37	−0.41
Kristen And Sara	128.19	128.11	34.40	34.21	92.66	92.46	2.89	4.51	0.73	−0.85	0.63	−0.34
	256.32	256.23	37.36	37.18	94.77	94.66	3.00	4.43	0.62	−0.53	0.44	−0.21
Four People	128.01	129.05	31.75	31.56	88.80	88.54	4.30	7.06	0.70	−0.34	1.23	−0.45
	256.05	257.70	34.94	34.59	92.73	92.51	4.35	6.33	0.72	−0.40	0.89	−0.36

Table 4. Control accuracy comparison of the reference and the proposed controller for intra frames using HM.13.

| Sequence	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR	ΔSSIM
	Ref	New	Ref	New	Ref	New	Ref	New	ROI (dB)	ROI
Johnny	640.00	639.99	28.78	28.90	83.61	83.70	2.60	2.90	0.35	0.94
	1280.04	1279.97	31.77	31.84	87.08	86.88	2.62	3.04	0.44	0.91
	2560.05	2559.93	34.84	35.00	91.44	91.00	2.41	2.88	0.74	1.04
Kristen And Sara	649.46	649.26	26.46	26.47	83.04	83.08	1.48	1.49	0.01	0.01
	1280.02	1280.07	29.31	29.44	87.27	87.17	1.21	1.31	0.40	0.65
	2560.02	2560.02	32.72	32.80	91.38	91.15	1.23	1.57	0.30	0.29
Four People	666.27	665.42	25.17	25.17	74.17	74.17	1.57	1.57	0.00	0.00
	1280.01	1279.88	27.10	27.09	78.99	78.80	1.40	1.32	−0.17	−0.35
	2559.98	2559.74	29.75	29.83	85.16	85.10	1.31	1.23	−0.19	−0.42
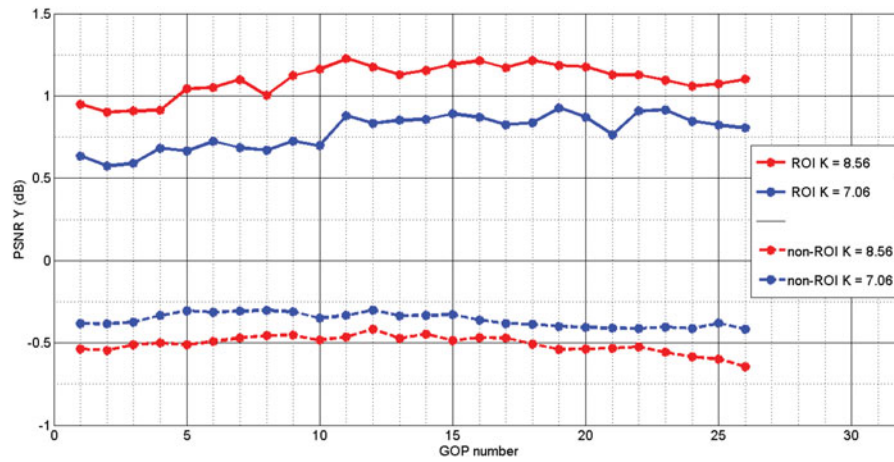
Fig. 6. ΔPSNR ROI and non-ROI (dB) for the last 25 GOPs of FourPeople at 128 kbps and using hierarchical bit allocation.

Table 5. Control accuracy comparison of the reference and the proposed controller for inter frames using HM.13.

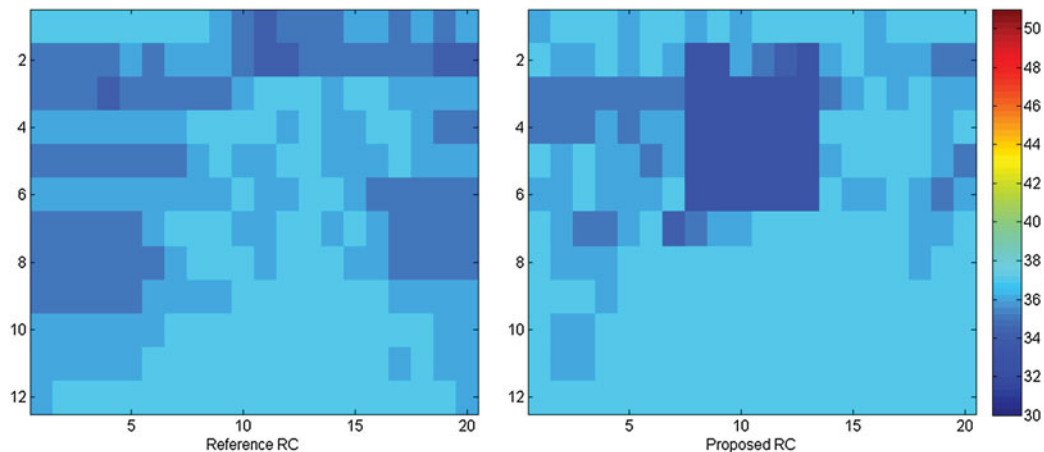| Sequence	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR(dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	Non-ROI	ROI	Non-ROI
Equal bit allocation												
Johnny	128.00	127.91	37.01	36.56	93.17	92.54	7.14	11.45	0.53	−0.69	0.39	−0.80
	256.01	255.82	39.49	39.01	95.20	94.75	6.57	11.01	0.53	−0.70	0.31	−0.56
Kristen And Sara	128.03	128.02	34.89	34.58	92.77	92.44	3.59	5.45	0.91	−0.59	0.81	−0.53
	256.05	256.01	37.75	37.46	94.84	94.60	3.15	5.21	0.89	−0.54	0.62	−0.38
Four People	128.03	128.03	32.36	32.13	90.07	89.83	4.71	7.51	0.84	−0.46	1.64	−0.48
	256.07	256.03	35.15	34.87	93.13	92.89	4.17	6.96	1.03	−0.53	1.32	−0.44
Hierarchical bit allocation												
Johnny	128.01	127.94	37.36	36.88	93.60	92.91	6.99	10.33	0.56	−0.75	0.43	−0.87
	256.01	256.32	39.74	39.26	95.40	94.98	6.94	10.72	0.55	−0.71	0.32	−0.54
Kristen And Sara	128.09	128.10	35.13	34.89	93.30	92.75	3.29	4.99	0.92	−0.50	0.79	−0.47
	256.10	256.08	37.91	37.65	95.01	94.80	3.19	4.92	0.92	−0.51	0.64	−0.35
Four People	128.02	128.27	32.58	32.35	90.42	90.15	5.16	7.34	0.88	−0.45	1.69	−0.52
	256.03	254.80	35.43	35.10	93.46	93.20	4.77	6.86	0.93	−0.55	1.14	−0.43
Adaptive bit allocation												
Johnny	128.00	127.87	37.48	37.00	93.74	93.05	6.53	9.93	0.54	−0.76	0.37	−0.86
	256.00	255.41	39.84	39.35	95.48	95.07	6.86	10.59	0.55	−0.73	0.33	−0.53
Kristen And Sara	128.05	128.09	35.21	34.96	93.14	92.84	3.19	4.87	0.94	−0.52	0.79	−0.48
	256.07	256.03	37.95	37.71	95.06	94.87	3.19	4.89	0.93	−0.50	0.63	−0.33
Four People	127.98	127.45	32.66	32.44	90.57	90.30	5.08	7.42	0.85	−0.43	1.47	−0.49
	255.98	253.94	35.50	35.17	93.55	93.28	4.85	6.93	0.90	−0.54	1.04	−0.43

Fig. 7. Comparison of QP repartition at the CTU level of Johnny.

Table 6. Control accuracy comparison of the reference and the proposed controller in HM.13

Sequence	Bit rate (kbps) Ref	New	PSNR Y (dB) Ref	New	SSIM Ref	New	K Ref	New	ΔPSNR(dB) ROI	Non-ROI	ΔSSIM ROI	Non-ROI
Johnny	128.00	128.02	35.95	36.11	92.59	92.25	5.50	9.06	1.59	−0.38	1.50	−0.56
	256.00	255.90	39.00	38.90	94.97	94.74	6.60	9.97	0.69	−0.33	0.30	−0.31
	512.01	511.34	41.09	40.84	96.16	96.86	6.58	10.71	0.48	−0.41	0.20	−0.24
	1500.01	1492.79	42.81	42.62	96.96	96.86	4.88	11.70	0.68	−0.35	0.26	−0.16
Kristen And Sara	129.86	128.18	33.21	33.72	91.76	91.93	2.71	4.30	1.91	−0.07	1.83	−0.11
	256.07	256.10	36.87	36.91	94.38	94.37	3.03	4.72	1.48	−0.34	1.06	−0.21
	512.07	512.00	39.76	39.60	95.97	95.89	3.03	4.76	0.95	−0.41	0.54	−0.18
	1500.10	1496.62	42.61	42.42	97.13	97.07	2.43	4.97	0.75	−0.39	0.34	−0.13
Four People	129.57	128.05	30.52	31.15	88.60	88.54	5.30	7.43	2.03	−0.22	3.18	−0.47
	256.00	255.48	34.29	34.26	92.64	92.33	5.02	7.10	1.46	−0.39	1.86	−0.59
	511.97	509.40	37.58	37.30	95.18	94.90	4.55	6.65	1.02	−0.55	0.92	−0.43
	1499.97	1484.96	41.46	41.18	97.05	96.93	3.87	6.70	0.78	−0.47	0.37	−0.18
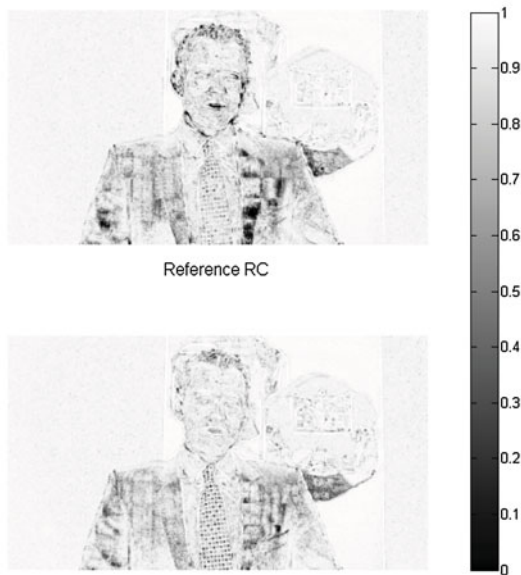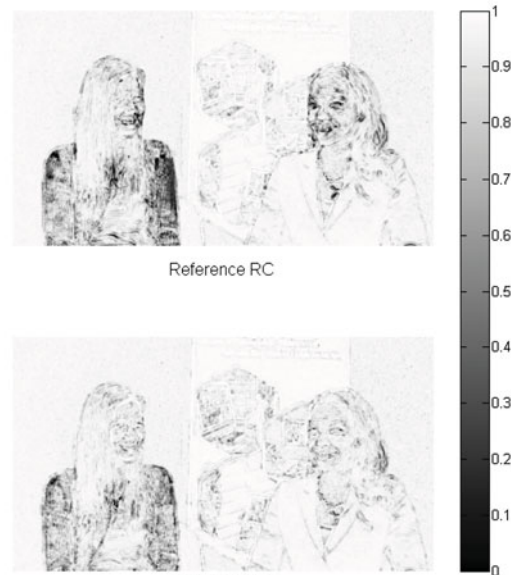
Fig. 8. Subjective comparison of Johnny coded at 128 kbps for an I frame. (a) Reference RC, (b) Proposed RC.

Fig. 9. Subjective comparison of Johnny coded at 128 kbps for a B frame. (a) Reference RC, (b) Proposed RC.

## 2. INTER PICTURE ROI-BASED ALGORITHM PERFORMANCE

A low delay B configuration is used to evaluate the performance of ROI-based allocation for B-frames. We first evaluate the global performance as done in HM.10. Results are given at 128 and 256 kbps to compare the performance with the first version of the controller. Equal, hierarchical, and adaptive bit allocations are tested.

From Table 5 we can deduce the same conclusions as in the previous version of our controller implemented in
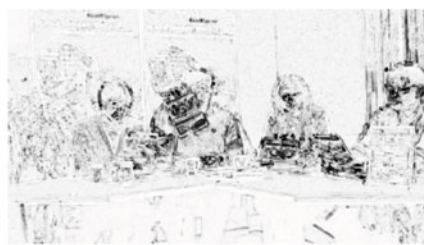
Fig. 10. Subjective comparison of KristenAndSara coded at 128 kbps for an I frame. (a) Reference RC, (b) proposed RC.

Fig. 11. Subjective comparison of KristenAndSara coded at 128 kbps for a B frame. (a) Reference RC, (b) proposed RC.
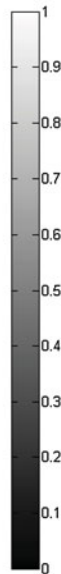

Fig. 12. Subjective comparison of FourPeople coded at 128 kbps for an I frame. (a) Reference RC, (b) proposed RC.

HM.10: the bit budget constraint is respected and ROI quality is improved proportionally to the repartition factor K.

At the CTU level the proposed approach gives a new QP distribution. Fig. 7 shows that smaller QP values are assigned to Johnny's face ($QP = 30$), while the rest of the frame takes bigger QPs that go from 34 to 38.

3. ROI-BASED ALGORITHM PERFORMANCE USING HYBRID CONFIGURATION

For a video conferencing system the low delay configuration is the most appropriate as we have the real-time constraint. However, to reduce packet loss effect and limit error propagation, an intra frame is introduced every second. Consequently, the final configuration of our encoder is the hybrid one. It handles a GOP of four B-frames coded in display order and an I-frame after 60 inter pictures. We choose the adaptive bit allocation at the frame level as it gives the best R-D performances and we tested four different rate points per sequence (128, 256, 512, and 1500 kbps).

From Table 6, we conclude that the controller global performance is maintained and the quality of the ROI is improved. At low bit rate, we can gain up to 2 dB in the ROI. Moreover, SSIM of the ROI is improved considerably when picture SSIM is smaller than 95. We can reach an improvement in the ROI quality of 3.18 dB for example. As SSIM is saturated when it gets close to 100, ΔSSIM is reduced when the picture index is higher than 95. We still in that case have noticeable improvement in ROI quality as the SSIM index goes from 0.20 to 0.92.
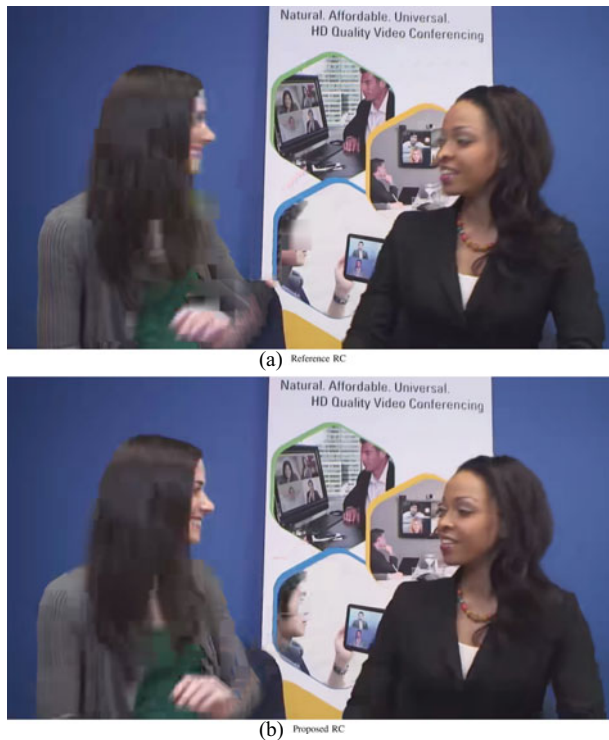
Fig. 13. Subjective comparison of FourPeople coded at 128 kbps for a B frame. (a) Reference RC, (b) proposed RC.
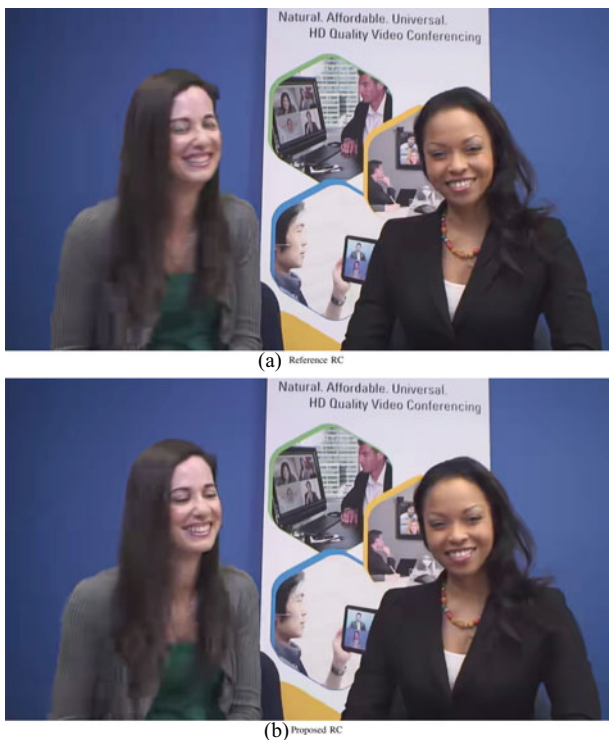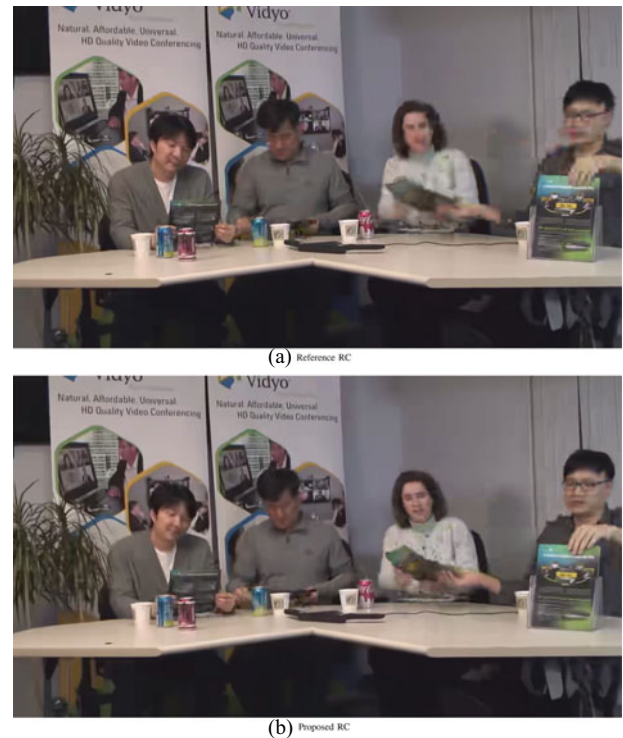

Fig. 15. SSIM map comparison KristenAndSara (a) Original Frame, (b) SSIM maps.

Fig. 14. SSIM map comparison Johnny. (a) Original frame, (b) SSIM maps.
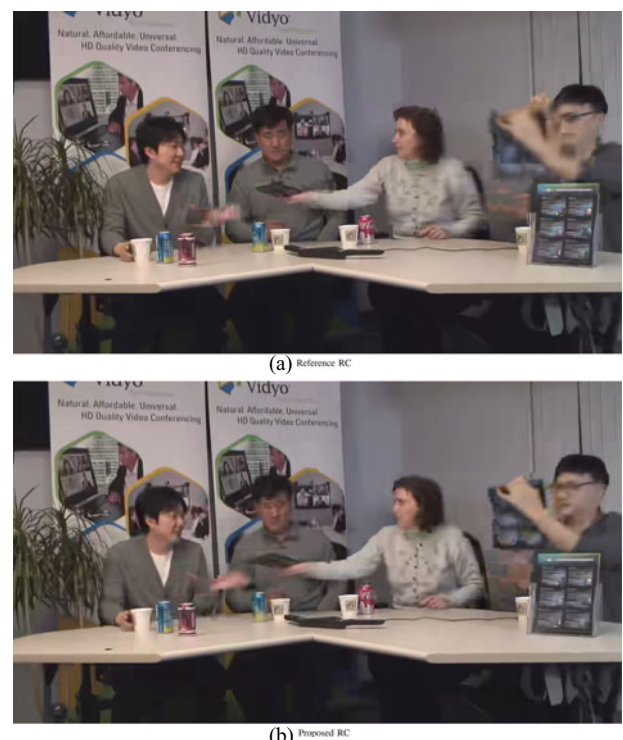

Fig. 16. SSIM map comparison FourPeople. (a) Original frame, (b) SSIM maps.

Experimental results show advantages in objective PSNR, in SSIM that predicts subjective opinion with high precision and visual evaluation for ROI as represented in Figs 8–13. We notice that for both intra and inter pictures and using our proposed scheme we can distinguish more details in the face and less artifacts, while the non-ROI does not present noticeable deterioration in visual quality as in video conferencing system the background is not changing in most of the cases.

Locally the SSIM index has been evaluated and an SSIM map has been computed for each frame to prove quality improvement in the ROI. Figs 14–16 represent the SSIM index over the whole frames (SSIM values goes from 0 for high distortion to 1 for high similarity). We notice that considering the proposed method SSIM index in the faces is closer to 1 (white faces). It shows an improvement in the details of the faces of the three tested sequences.

E) Comparison of the proposed R-λ ROI-based controller with URQ ROI-based controller

The last step consists in using an ROI-based RC algorithm initially proposed for H.264 and based on the quadratic model, then, adapt it to HEVC as described in Section IV. The performed tests in this section use a low delay
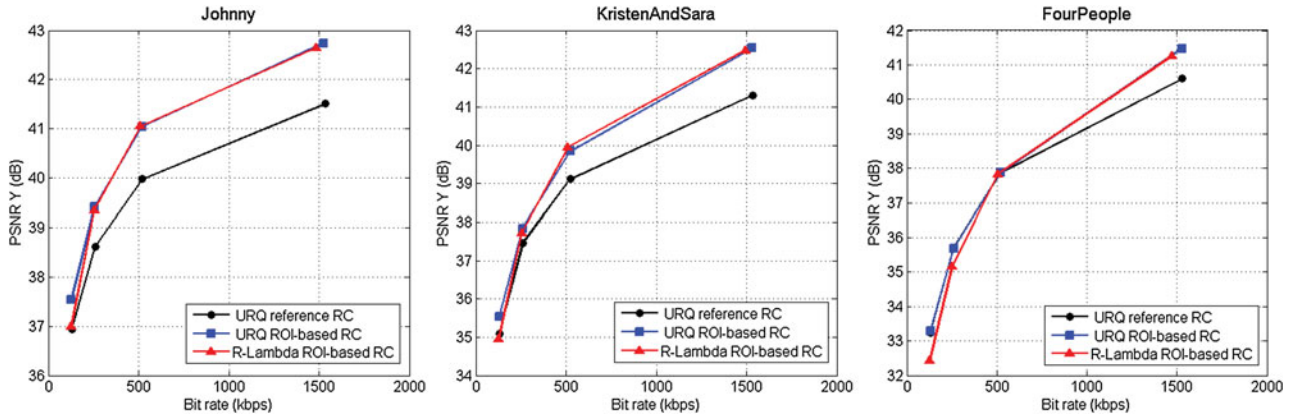

Fig. 17. R-D performances of R-λ ROI-based algorithm and URQ ROI-based model compared to URQ reference RC algorithm.

Table 7. RC results using URQ model at 128 kbps.

| Sequence		K	Bit rate (kbps)	PSNR Y (dB)	SSIM	PSNR ROI (dB)	SSIM ROI	PSNR non-ROI (dB)	SSIM non-ROI
Johnny	Ref	3.93	130.48	36.94	93.66	32.60	94.18	38.28	93.59
	New	7.90	129.19	37.54	93.59	35.25	95.21	38.04	93.39
		8.54	129.08	37.56	93.59	35.45	95.25	38.00	93.38
Kristen And Sara	Ref	2.13	130.92	35.10	93.10	31.74	93.75	36.04	93.02
	New	4.26	130.73	35.54	93.24	34.02	94.35	35.85	93.10
		4.59	130.65	35.60	93.30	34.29	94.49	35.87	93.15
Four People	Ref	4.31	129.87	33.25	91.38	29.92	82.86	33.94	92.46
	New	6.08	129.69	33.30	91.38	30.68	84.30	33.80	92.28
		6.42	129.94	33.28	91.36	30.87	84.67	33.73	92.21
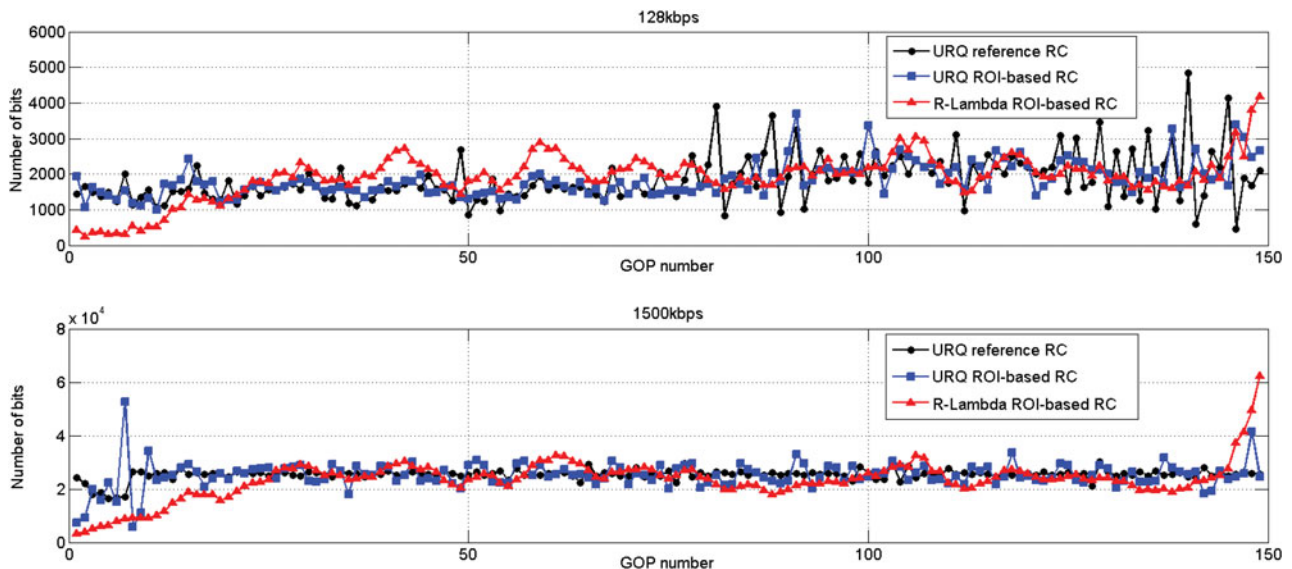
Fig. 18. Comparison of bit fluctuation per GOP of R-λ and URQ ROI-based models at low and high bit rate for sequence Johnny.

configuration where all frames are coded in bidirectional mode (B-frames). We tested the three sequences at four different bit rates (128, 256, 512, and 1500 kbps).

We notice from Table 7 that the URQ ROI-based method implemented in HM.9 respects the budget constraint at low bit rate. It is also the case at high bit rate as represented in Fig. 17.

R–D performance evaluation shows an important improvement in rate control performances. The obtained R-D curve
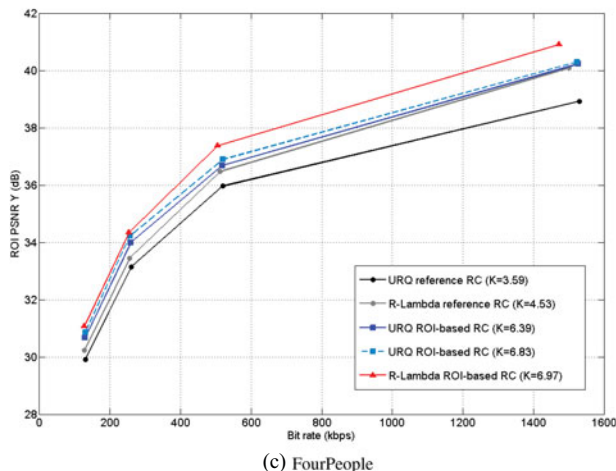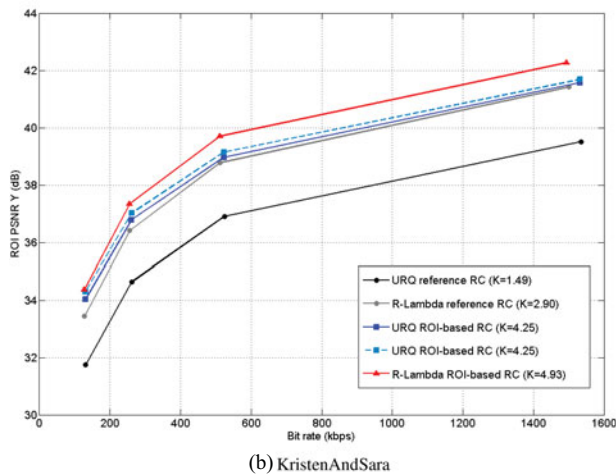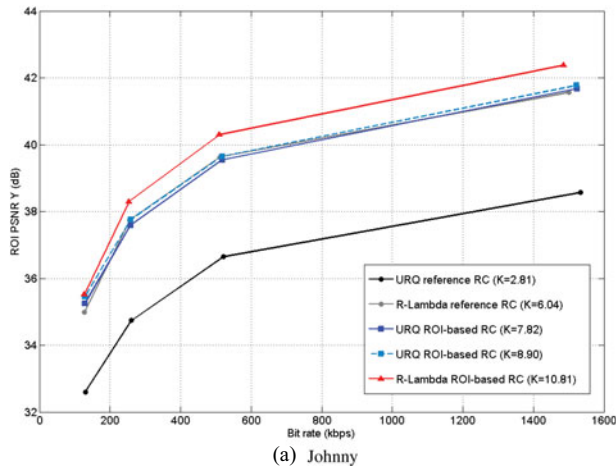
is better than the reference URQ model and comparable to the one given by our R-λ algorithm implemented in HM.13. Moreover, table 7 shows that URQ ROI-based method improves the quality of the ROI while using higher bit ratio K.

Fig. 18 shows bit distribution over GOP at low and high bit rates for Johnny sequence. We conclude that the proposed R-λ method gives a smoother bit allocation compared to the URQ methods at low bit rate with no unsettled bit picks, while at high bit rate the three algorithms gives comparable distribution over GOPs. The same conclusion is valid for all tested sequences.

Fig. 19 represents R-D performance of all evaluated methods. It gives the overall ROI PSNR for each bit rate. For the three tested sequences, the reference URQ controller has the worst R–D performances. Once introducing the ROI, both URQ-based method and R-λ-based method show better R-D performance compared with the reference.

Finally, in the URQ scheme ROI-based bit allocation is only performed for referenced frames of type B. Our algorithm (based on R-λ method) makes ROI-based allocation for all frame types, which leads to a better QP repartition over regions in the full sequence. With our algorithm we can reach higher ratios K, as shown in Fig. 19.

VI. CONCLUSION

In this paper, an ROI-based rate control for HEVC is proposed for HM.10 and improved for HM.13. A second ROI-based RC algorithm studied in the state-of-the-art has been adapted to HEVC controller and implemented in HM.9.

Our proposed scheme uses the R-λ model, takes into account both I and B frames and achieves better visual quality in ROIs thanks to an independent processing of ROI and non-ROI regions at the CTU level and a larger QP clipping range, while the reference scheme used the quadratic model and performs rate control only in referenced B-frames.

The proposed algorithms lead to better quality in ROI, while respecting the global bit rate constraint. All implemented schemes are useful for video conferencing systems to allow a better representation of the facial expression.

Fig. 19. Comparative ROI-based R-D performances of different methods. (a) Johnny, (b) KristenAndSara, and (c) FourPeople.

REFERENCES

[1] Wu, Z.; Xie, S.; Zhang, K.; Wu, R.: Rate control in video coding, in Recent Advances on Video Coding, J. Del Ser Lorente, Ed. *InTech*, 2011, 79–117. http://www.intechopen.com/books/recent-advances-on-video-coding/rate-control-in-video-coding

[2] Chen, Z.; Ngan, K.N.: Recent advances in rate control for video coding. Signal Process., Image Commun., **22** (1) (2007), 19–38.

[3] Yang, L.; Zhang, L.; Ma, S.; Zhao, D.: A ROI quality adjustable rate control scheme for low bitrate video coding, in *Proc. Picture Coding Symp.*, IEEE, May 2009, 1–4.

[4] Chiang, J.; Hsieh, C.; Chang, G.; Jou, F.-D.; Lie, W.-N.: Region-of-interest based rate control scheme with flexible quality on demand, in *Proc. IEEE Int. Conf. on Multimedia and Expo*, 2010, pp. 238–242.

[5] C.-Wu, Y.; P.-Su, C.: A region-of-interest rate-control scheme for encoding traffic surveillance videos, in *Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing*, 2009, pp. 194–197.

[6] Hu, H.; Li, B.; Lin, W.; Li, W.; Sun, M.-T.: Region-based rate control for H.264/AVC for low bit-rate applications. IEEE Trans. Circuits Syst. Video Technol., **22** (11) (2012), 1564–1576.

[7] Sullivan, G.J.; Ohm, J.-R.; Han, W.-J.; Wiegand, T.: Overview of the high efficiency video coding (hevc) standard. IEEE Trans. Circuits Syst. Video Technol., **22** (12) (2012), 1649–1668.

[8] Choi, H.; Nam, J.; Yoo, J.; Sim, D.: Rate control based on unified RQ model for HEVC, in *JCT-VC H0213, 8th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, San José, CA, USA, 2012.

[9] Choi, H.; Nam, J.; Yoo, J.; Sim, D.: Improvement of the rate control based on pixel-based URQ model for HEVC, in *JCT-VC I0094, 9th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2012.

[10] Li, B.; Li, H.; Li, L.; Zhang, J.: Rate control by R-lambda model for HEVC, in *JCT-VC K0103, 11th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Shanghai, CN, 2012.

[11] Li, B.; Li, H.; Li, L.: Adaptive bit allocation for R-lambda model rate control in HM, in *JCT-VC M0036, 13th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Incheon, KR, 2013.

[12] Karczewicz, M.; Wang, X.: Intra frame rate control based SATD, in *JCT-VC M0257, 13th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Incheon, KR, 2013.

[13] Lee, B.; Kim, M.; Nguyen, T.: A frame-level rate control scheme based on texture and non-texture rate models for high efficiency video coding. IEEE Trans. Circuits Syst. Video Technol. **24** (2013), 1–14.

[14] Chuang, T.-H.; Cheng, K.-Y.; Yung-Yu, H.: A collaborative benchmark for region of interest detection algorithms, in *IEEE Conf. on Computer Vision and Pattern Recognition*, June 2009, 296–303.

[15] Zahir, N.B., Samad, R.; Mustafa, M.: Initial experimental results of real-time variant pose face detection and tracking system, in *IEEE Int. Conf. on Signal and Image Processing Applications*, October 2013, 264–268.

[16] Viola, P.; Jones, M.: Rapid object detection using a boosted cascade of simple features, in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. CVPR*, **1** (2001), I-511–I-518.

[17] Li, Q.; Niaz, U.; Merialdo, B.: An improved algorithm on Viola-Jones object detector, in *Content-Based Multimedia Indexing (CBMI)*. IEEE, June 2012, 1–6.

[18] Sullivan, G.J.; Wiegand, T.: Rate-distortion optimization for video compression. IEEE Signal Process. Mag., **15** (1998), 74–90.

[19] Wiegand, T.; Schwarz, H.; Joch, A.; Kossentini, F.; Sullivan, G.J.: Rate-constrained coder control and comparison of video coding standards. IEEE Trans. Circuits Syst. Video Technol., **13** (7) (2003), 688–703.

[20] Ma, S.; Si, J.; Wang, S.: A study on the rate distortion modeling for high efficiency video coding, in *Proc. of IEEE Int. Conf. Image Proc.*, September 2012, 181–184.

[21] Yu, S.; Ahmad, I.: A new rate control algorithm for MPEG-4 Video Coding, in *Proc. SPIE*, 2002.

[22] Li, Z.; Gao, W.; Pan, F.; Ma, S.; Lim, K.; Feng, G. et al.: Adaptive rate control for h.264. Elsevier J. Vis. Commun. Image Repres., **17** (2) (2006), 376–406.

[23] Naccari, M.; Pereira, F.: Quadratic modeling rate control in the emerging HEVC standard, in *Proc. Picture Coding Symp.*, 2012, pp. 401–404.

[24] Sun, L.; Au, O.; Dai, W.: An adaptive frame complexity based rate quantization model for intra-frame rate control of high efficiency video coding (HEVC). *APSIPA ASC*, 2012, 1–6.

[25] Si, J.; Ma, S.; Gao, W.: Adaptive rate control for HEVC, in *JCT-VC I0433, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2012.

[26] Yoon, Y.; Kim, H.; Jung, S.-h., Jun, D.: A new rate control method for hierarchical video coding in HEVC, in *IEEE Int. Symp. on Broadband Multimedia Systems and Broadcasting*, 2012.

[27] He, Z.; Mitra, S.: Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling. IEEE Trans. Circuits Syst. Video Technol., **12** (10) (2002), 840–849.

[28] Ma, Z.; Xu, M.; Ou, Y.; Wang, Y.: Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications. IEEE Trans. Circuits Syst. Video Technol., **22** (5) (2012), 671–682.

[29] Ma, S.; Gao, W.; Lu, Y.: Rate-distortion analysis for H.264/AVC video coding and its application to rate control. IEEE Trans. Circuits Syst. Video Technol., **15** (12) (2005), 1533–1544.

[30] Bossen, F.: Common test conditions and software reference configurations, in *JCT-VC L1100, 12th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013.

[31] HEVC test model 13 (HM.13). [Online]: Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-13.0/

[32] HEVC test model 9 (HM.9). [Online]: Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-9.0/

[33] HEVC test model 10 (HM.10). [Online]: Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-10.1/

[34] Kim, I.-K.; McCann, K.; Sugimoto, K.; Bross, B.; Han, W.-j., Sullivan, G.: High Efficiency Video Coding (HEVC) Test Model 13 (HM13) Encoder Description, in *JCT-VC O1002, 15th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013.

[35] Zhao, T.; Zeng, K.; Rehman, A.; Wang, Z.: On the use of SSIM in HEVC, in *Asilomar Conf. on Signals, Systems and Computers*. Pacific Grove, California, USA: IEEE, November 2013, pp. 1107–1111.

Marwa Meddeb (S′13) received the double engineering degree in telecommunications from the High School of Communication of Tunis (SUP'COM), Tunis, Tunisia and Télécom ParisTech, Paris, France, in 2013 thanks to a double degree program in telecommunications and multimedia technologies within EURECOM, Sophia Antipolis, France. Since February 2013, she is working toward the Ph.D. degree within the Image and Signal Processing Department at Télécom ParisTech (Paris, France). Her research interests includes image and video coding, segmentation and transmission.

Marco Cagnazzo (M′05-SM′11) obtained the Laurea (equivalent to the M.S.) degree in Telecommunication Engineering

from Federico II University, Napoli, Italy, in 2002, and the Ph.D. degree in Information and Communication Technology from Federico II University and the University of Nice-Sophia Antipolis, Nice, France in 2005.

He was a post-doc fellow at I3S Laboratory (Sophia Antipolis, France) from 2006 to 2008. Since February 2008 he has been Associate Professor at Institut Mines-Télécom, Télécom ParisTech (Paris), within the Multimedia team. He is author of more than 90 contributions in peer-reviewed journals, conferences proceedings, books and book chapters. His current research interests are three-dimensional video communication and coding, distribued video coding, robust video delivery, network coding.

Dr. Cagnazzo is an Area Editor for *Elsevier Signal Processing: Image Communication* and *Elsevier Signal Processing*. Moreover he is a reviewer for major international scientific reviews (IEEE Trans. Multimedia, IEEE Trans. Image Processing, IEEE Trans. Signal Processing, IEEE Trans. Circ. Syst. Video Tech., *Elsevier Signal Processing*, *Elsevier Sig. Proc. Image Comm.*, and others) and conferences (IEEE International Conference on Image Processing, IEEE MMSP, European Signal Processing Conference, and others).

Béatrice Pesquet-Popescu (SM′06, F′13) received the engineering degree in telecommunications from the "Politehnica" Institute in Bucharest in 1995 (highest honours) and the Ph.D. degree from the Ecole Normale Supérieure de Cachan in 1998. In 1998, she was a Research and Teaching Assistant with Université Paris XI, Paris. In 1999, she joined Philips Research France, Suresnes, France, where she worked for two years as a Research Scientist, then as a Project Leader, in scalable video coding. Since Oct. 2000 she is with Télécom ParisTech (formerly, ENST), first as an Associate Professor, and since 2007 as a Professor, Head of the Multimedia Group. She is the Head of the UBIMEDIA common research laboratory between Alcatel-Lucent and Institut Télécom. Her current research interests are in source coding, scalable, robust and distributed video compression and sparse representations. Dr. Pesquet-Popescu was an EURASIP BoG member (2003–2010), and an IEEE Signal Processing Society IVMSP TC member and MMSP TC associate member. She serves as an Associate Editor for IEEE Trans. on Image Processing, IEEE Trans. on Multimedia, IEEE Trans. on CSVT, Elsevier Image Communication, and Hindawi Int. J. Digital Multimedia Broadcasting journals and was till 2010 an Associate Editor for Elsevier Signal Processing. She was a Technical Co-Chair for the PCS2004 conference, and General Co-Chair for IEEE SPS MMSP2010, EUSIPCO 2012, and IEEE SPS ICIP 2014 conferences. Beatrice Pesquet-Popescu is a recipient of the "Best Student Paper Award" in the IEEE Signal Processing Workshop on Higher-Order Statistics in 1997, of the Bronze Inventor Medal from Philips Research and in 1998 she received a "Young Investigator Award" granted by the French Physical Society. She holds 23 patents in wavelet-based video coding and has authored more than 290 book chapters, journal and conference papers in the field. In 2006, she was the recipient, together with D. Turaga and M. van der Schaar, of the IEEE Trans. on Circuits and Systems for Video Technology "Best Paper Award".