SUBJECTIVE-QUALITY-OPTIMIZED COMPLEXITY CONTROL FOR HEVC DECODING

Ren Yang, Mai Xu*, Lai Jiang, Zulin Wang

The School of Electronic and Information Engineering, Beihang University, China
{yangren, maixu, jianglai.china, wzulin}@buaa.edu.cn

ABSTRACT

The latest High Efficiency Video Coding (HEVC) standard significantly improves coding efficiency over H.264/AVC, at the cost of heavy encoding and decoding complexity. For reducing HEVC decoding complexity to a target, we propose in this paper a Subjective-Quality-Optimized Complexity Control (SQOCC) approach, which optimizes subjective quality loss caused by the decoding complexity reduction. First, a saliency detection method in HEVC domain is developed as the preliminary of subjective quality metric. Based on detected saliency, we establish a formulation to minimize subjective quality loss at the constraint of specific decoding complexity reduction, via disabling the deblocking filters of some Largest Coding Units (LCUs). Next, we utilize least square fitting to model functions in our formulation. We then provide a solution to our formulation, achieving subjective-quality-optimized complexity control for HEVC decoding. Finally, the experimental results show the effectiveness of our SQOCC approach in terms of both control accuracy and subjective quality.

Index Terms— HEVC, complexity control, saliency, subjective quality, deblocking filter.

1. INTRODUCTION

The past decade has witnessed an explosive growth of video data delivered over Internet. Such a growth poses a great challenge on the efficiency of video coding. To significantly improve the efficiency of video coding, the High Efficiency Video Coding (HEVC) standard [1] was approved in April, 2013. It has been verified that HEVC can reduce the bit-rate to around 50% with similar visual quality, compared with its former standard H.264/AVC. However, the cost for the high coding efficiency is much more computational complexity [2]. This cost becomes a great obstacle of video encoding and decoding on portable devices with limited power.

There are extensive approaches on accelerating the decoding speed with hardware techniques, such as [3, 4, 5, 6, 7]. For example, Yan *et al.* [5] and Chi *et al.* [6] proposed to take advantages of Single Instruction Multiple Data (SIMD) instructions for increasing decoding speed. Souza *et al.* [7]

Fig. 1. An example for application of our SQOCC approach. When the devices is with sufficient power (eg. 80%), the video can be decoded with full fidelity. Once the battery is insufficient (eg. 77%, 73% and 68%), the video is decoded with low complexity (complexity is reduced in the dark areas), according to the left battery. However, it incurs the quality degradation. The subjective quality should be maximized in quality degradation for the proper Quality of Experience (QoE).

achieved the decoding acceleration by proposing a novel GPU parallel algorithm. The aforementioned approaches can only accelerate decoding, but cannot reduce the complexity and power consumption. In addition, [8] firstly predicts the decoding computational complexity of each frame, and then changes the frequency of CPU to the minimum frequency on the promise of real-time decoding. However, all above approaches can only be applied in specific hardware at the decoder side. They are hardly adaptive to all general hardware.

To overcome the drawback of hardware-based approaches, some algorithmic approaches have been proposed to decrease video decoding complexity. There exist a great number of works, such as [9, 10, 11, 12, 13]. For H.264/AVC, Liu *et al.* [9] proposed to detect Region-Of-Interest (ROI), and to allocate less computational resources to non-ROIs. As such, the total encoding and decoding complexity can be reduced by an ROI based Rate-Distortion-Complexity (R-D-C) cost function. Later, Naccari *et al.* [10] proposed an approach for reducing decoding complexity of H.264/AVC and HEVC. The approach optimized by the Generalized Block-edge Impairment Metric (GBIM) estimates the offsets in the deblocking filters, with lower complexity than the conventional brute force optimization. For HEVC, [11] reduces the decoding complexity by modifying the structure of prediction during encoding. Later, [12] and [13] proposed to remove the HEVC in-loop filters and to shorten the sizes of the FIR filters in motion compensation, for reducing HEVC decoding complexity. Nevertheless, the above approaches cannot control the decoding complexity to a given target, leading to insufficient or wasteful use of power resources.

There are few works on controlling the decoding complexity. To be more specific, Langroodi *et al.* [14] developed

an approach that the decoder sends its computational resource demand to the encoder side, and the encoder optimizes the motion compensation to achieve decoding complexity control at the decoder side. However, the configuration of encoder has to be changed according to the feedback. Thus, their approach dose not work in the case that videos have been already encoded. Besides, it does not aim at optimizing perceptual quality, which is a key for Visual Quality Assessment (VQA).

In this paper, we propose a Subjective-Quality-Optimized Complexity Control (SQOCC) approach to control the HEVC decoding complexity with minimal subjective quality loss. Fig. 1 shows a possible application of our approach. The basic idea of our approach is to disable the deblocking filters of some non-salient Largest Coding Units (LCUs), according to the target decoding complexity. The computational complexity can be reduced when disabling the deblocking filters of some LCUs. The cost is the degraded visual quality. Accordingly, there are two objectives of our SQOCC approach: 1) Reducing the decoding complexity to the target, via disabling the deblocking filters of some LCUs; 2) Optimizing the subjective quality loss, via considering the detected saliency of each LCU in disabling deblocking filters. The main contributions of our approach are:

- We propose an optimization formulation to minimize subjective quality loss at a given target complexity of HEVC decoding.
- We model the relationship among saliency, subjective quality loss and complexity reduction, when disabling debocking filters of some LCUs in HEVC decoding.
- We develop a solution to the proposed formulation, achieving complexity control of HEVC decoding with minimal subjective quality distortion.

2. SALIENCY DETECTION

2.1. The proposed method

When reducing decoding complexity of HEVC, the visual quality may be degraded as the expense. In fact, subjective quality can be favored in visual quality degradation, as the Human Visual System (HVS) [15] normally pays attention to small salient regions. Saliency detection [16] aims at predicting visual attention of humans, and it can be used for favoring subjective quality during decoding complexity control. Most existing saliency detection methods [16] work in pixel domain. As such, they cannot be used in decoding complexity control, since pixels are not available before decoding videos but required for saliency detection. Most recently, Operational Block Description Length (OBDL) has been exploited in [17], which is based on the bit number of each coding block assigned in the H.264 encoder. Then, the OBDL is embedded into Markov Random Fields (MRF) for video saliency detection. However, the energy optimization of MRF makes OBDL too time-consuming to be implemented in decoding complexity control. Besides, OBDL can only be used in H.264 bit-

streams, rather than HEVC. Thus, we develop a new saliency detection method in HEVC domain as the preliminary for our SQOCC approach.

In [16], it has been argued that the region with high-information can attract more visual attention. The HEVC encoder also prefers to assign more bits to high-information regions. Therefore, the number of allocated bits of each LCU, denoted by b_n for the n-th LCU, is seen as a feature in our saliency detection method. Moreover, [16] has also stated that contrast can be used for saliency detection, since the region standing out from its neighbors may attract extensive attention. So, the contrast of bit allocation is considered as another saliency detection feature in our method. Specifically, Δb_n, the contrast of bit allocation of the n-th LCU, can be calculated as

$$\Delta b_n = \left(\frac{\sum_{n' \in \mathbf{I}} \exp(-\frac{d_{n'}^2}{\sigma_b^2})(b_{n'} - b_n)^2}{\sum_{n' \in \mathbf{I}} \exp(-\frac{d_{n'}^2}{\sigma_b^2})} \right)^{\frac{1}{2}}, \qquad (1)$$

where I is the set of 8-neighboring LCUs and $d_{n'}$ is the Euclidean distance between the n'-th and n-th LCUs. In addition, σ_b is a parameter to control the spatial contrast of bit allocation. In our saliency detection method, we set $\sigma_b = 1.2$ to make the results appropriate. Then, both b_n and Δb_n are normalized and linearly combined as follows,

$$w_n = \frac{1}{2}(\frac{b_n}{b_{max}} + \frac{\Delta b_n}{\Delta b_{max}}), \qquad (2)$$

where b_{max} and Δb_{max} are the maximal b_n and Δb_n in the video frame. Finally, the saliency map of each video frame can be obtained[1] for our SQOCC approach.

2.2. Effectiveness validation

Now, we evaluate the performance of our saliency detection method in HEVC domain. The performance is evaluated over all 15 video sequences of classes B, C and E/E′ from JCT-VC database [18]. For performance evaluation, the saliency detection accuracy is measured in terms of the Area Under ROC Curve (AUC), Normalized Scanpath Saliency (NSS) and linear Correlation Coefficient (CC). Here, the accuracy of our method is averaged over all sequences, and then it is compared with that of some state-of-the-art saliency detection methods, i.e., PQFT [19], Rudoy *et al.* [20], OBDL [17] and Itti's model [21]. The results are tabulated in Table 1. Note that the method with a larger AUC, NSS, or CC value can better predict the human fixations. We can see from this table that our saliency detection method is comparable to or even better than other methods. This validates the effectiveness of our method in saliency detection.

3. COMPLEXITY CONTROL APPROACH

3.1. Formulation for SQOCC approach

Our SQOCC approach aims at reducing decoding complexity to the target, meanwhile favoring subjective quality. Here,

---

[1]Saliency detection consumes averagely 0.2ms (1.2×10^{-3}mWh) per frame.

Table 1. Comparison of saliency detection accuracy

	Our	PQFT	Rudoy	OBDL	Itti
AUC	0.80	0.72	0.77	0.79	0.70
NSS	1.27	0.88	1.32	1.26	0.49
CC	0.36	0.23	0.37	0.32	0.12

ΔS_n and ΔC_n are defined as the subjective quality loss and complexity reduction of the n-th LCU in a frame, and ΔC_T is the target of complexity reduction. The optimization of SQOCC can be expressed by the following formulation:

$$\min \sum_{n=1}^{N} \Delta S_n \quad \text{s.t.} \quad \sum_{n=1}^{N} \Delta C_n \geq \Delta C_T, \tag{3}$$

where N is the total number of LCUs in a frame.

In fact, [2] has verified that deblocking filters consume 13%-17% of decoding complexity. Hence, the decoding complexity can be reduced by disabling the deblocking filters of some LCUs, at the expense of visual quality loss. On the other hand, the subjective quality loss should be minimized according to the HVS. $f_n \in \{0, 1\}$ indicates whether the deblocking filter of the n-th LCU is disabled ($f_n = 0$) or enabled ($f_n = 1$). Through experimental analysis, we found that when the deblocking filter is disabled, ΔS_n and ΔC_n in (3) have strong correlation with saliency value of the n-th LCU. Thus, formulation (3) can be turned to

$$\min_{\{f_n\}_{n=1}^{N}} \sum_{n=1}^{N} \Delta S(f_n, w_n) \quad \text{s.t.} \quad \sum_{n=1}^{N} \Delta C(f_n, w_n) \geq \Delta C_T, \tag{4}$$

where $\Delta S(f_n, w_n)$ indicates a function of subjective quality loss regarding f_n and w_n, and $\Delta C(f_n, w_n)$ is a function of complexity reduction with respect to f_n and w_n. Next, we focus on modelling $\Delta S(f_n, w_n)$ and $\Delta C(f_n, w_n)$ for our SQOCC approach.

3.2. Relationship Modelling for $\Delta S(f_n, w_n)$

According to [22], subjective visual distortion can be measured by weighted Mean Square Error (MSE). Assume that w_n and MSE_n are the saliency weight and MSE of the n-th LCU, respectively. Then, subjective quality loss $\Delta S(f_n, w_n)$ can be calculated by

$$\Delta S(f_n, w_n) = \frac{w_n \text{MSE}_n}{\sum_{n=1}^{N} w_n}(1 - f_n) = \frac{1}{N}\frac{w_n \text{MSE}_n}{\overline{w}}(1 - f_n), \tag{5}$$

where $\overline{w} = \sum_{n=1}^{N} w_n/N$ is the saliency weight averaged over all LCUs. Since we can see from (5) that $\Delta S(f_n = 1, w_n) = 0$, we only focus on training function $\Delta S(f_n = 0, w_n)$ in the following way.

Training Sequences. The training sequences were randomly selected from JCT-VC database [18], including two 1920×1080 sequences *Cactus* and *BasketballDrive* from Class B, two 832×480 sequences *BQMall* and *BasketballDrill* from Class C, and three 1280×720 sequences *KristenAndSara*, *Vidyo1* and *Vidyo3* from Class E/E'. We empirically found that both video content and bit rate have effect on the model of $\Delta S(f_n = 0, w_n)$. Thus, according to the video content, we divided the sequences into two training sets. The
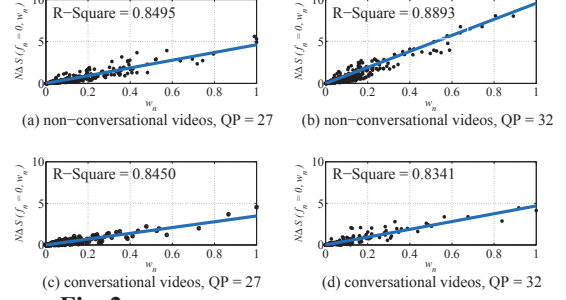
Fig. 2. Fitting Curves of w_n and $N\Delta S(f_n = 0, w_n)$.

Table 2. The values of l, a and b at different scenarios

	non-conversational		conversational	
	QP = 27	QP = 32	QP = 27	QP = 32
l	4.613	9.520	3.484	4.794
a	0.3724	0.4655	0.4089	0.3633
b	0.0520	0.0636	0.0686	0.0669

first set consists of non-conversational videos, i.e., the four sequences from Classes B and C. The second set is comprised by the three conversational videos from Class E/E'. The sequences in each set were compressed by HM 16.0 at two different Quantization Parameters (QPs), i.e., QP = 27 and 32. All parameters in HM 16.0 are the same as those in Section 4.

Training Procedure. We decoded the training sequences with deblocking filter being enabled and disabled, respectively. As such, the MSE_n can be obtained as the visual quality loss, when disabling deblocking filter for each LCU. Then, we randomly selected 300 LCUs in each scenario[2], and recorded their corresponding w_n and MSE_n.

Training Results. Given w_n and MSE_n of each randomly selected LCU, its $N\Delta S(f_n = 0, w_n)$ can be achieved by (5). We apply the first-order least square fitting on $(w_n, N\Delta S(f_n = 0, w_n))$ of all selected 300 LCUs, and the fitting curves are shown in Fig. 2. As a result, the fitting equation is obtained as follows,

$$N\Delta S(f_n = 0, w_n) = l \cdot w_n, \tag{6}$$

where the values of l at different training sets of sequences and QPs are reported in Table 2. Finally, $\Delta S(f_n, w_n)$ was achieved upon (6). Note that for other QPs, the relationship should be retrained.

3.3. Relationship Modelling for $\Delta C(f_n, w_n)$

For modelling $\Delta C(f_n, w_n)$, we define $C_d(w_n)$ as the complexity of deblocking filter at the n-th LCU, which is also related to its corresponding saliency weight. C_{Total} is the decoding complexity of a frame with deblocking filters enabled. Then, $\Delta C(f_n, w_n)$ can be calculated by

$$\Delta C(f_n, w_n) = \frac{C_d(w_n)}{C_{\text{Total}}}(1 - f_n) = \frac{1}{N}\frac{C_d(w_n)}{\overline{C_{\text{LCU}}}}(1 - f_n), \tag{7}$$

where $\overline{C_{\text{LCU}}} = C_{\text{Total}}/N$ is the average decoding complexity for each LCU. Also, we use the training sequences of Section 3.2 to model $\Delta C(f_n = 0, w_n)$. In the following, we briefly

---

[2]There are four scenarios including two sets of non-conversational and conversational sequences at two QPs.

(a) non-conversational videos, QP = 27   (b) non-conversational videos, QP = 32

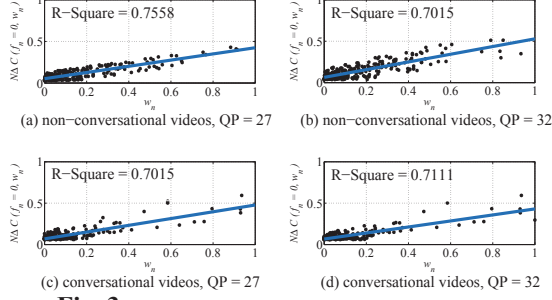(c) conversational videos, QP = 27   (d) conversational videos, QP = 32

Fig. 3. Fitting Curves of w_n and $N\Delta C(f_n = 0, w_n)$.

present the procedure and results of the training on $\Delta C(f_n = 0, w_n)$.

Training Procedure. The decoding complexity was measured by the software Intel® Power Gadget 3.0, on a Windows PC with Inter(R) Core(TM) i7-4790K CPU. First, all training sequences (compressed by HM 16.0 at QP = 27 and 32) were decoded with deblocking filters enabled. In this case, w_n and $C_n(w_n)$ of each LCU can be recorded. Then, we randomly selected w_n and $C_n(w_n)$ of 300 LCUs from each scenario to model $\Delta C(f_n = 0, w_n)$.

Training Results. Again, we apply the first-order least square fitting on $(w_n, N\Delta C(f_n = 0, w_n))$ of all selected LCUs. The fitting curves are plotted in Fig. 3. Consequently, the function of $N\Delta C(f_n = 0, w_n)$ is obtained as follows,

$$N\Delta C(f_n = 0, w_n) = a \cdot w_n + b, \qquad (8)$$

where the values of a and b at different sets and QPs are presented in Table 2. Finally, $\Delta C(f_n, w_n)$ can be modelled.

3.4. Solution to SQOCC optimization formulation

Now, based on (6) and (8), formulation (4) for our SQOCC approach can be rewritten in the following,

$$\min_{\{f_n\}_{n=1}^N} \sum_{n=1}^N \frac{1}{N} l \cdot w_n \cdot (1 - f_n)$$

$$\text{s.t.} \quad \sum_{n=1}^N \frac{1}{N}(a \cdot w_n + b)(1 - f_n) \geq \Delta C_T. \qquad (9)$$

Then, we concentrate on finding optimal solution set $\mathbf{F} = \{f_n\}_{n=1}^N$ towards (9). First, let $\{\widetilde{w}_n\}_{n=1}^N$ be the set of the ascending sorted[3] $\{w_n\}_{n=1}^N$. Given $\{\widetilde{w}_n\}_{n=1}^N$, Lemma 1 can be used for finding the optimal solution to (9).

Lemma 1: Let $a > 0$, $b > 0$, $l > 0$ and $w_n \in [0, 1]$. Assume that $\mathbf{F} = \{f_n\}_{n=1}^N$ satisfies

$$f_n = \begin{cases} 0, & w_n \leq \widetilde{w}_I \\ 1, & \text{otherwise}, \end{cases} \qquad (10)$$

where $\widetilde{w}_I$ is the I-th value of ascending sorted $\{w_n\}_{n=1}^N$. Assume that $\mathbf{F}' = \{f_n'\}_{n=1}^N$ is another set, where $f_n' \in \{0, 1\}$. If

$$\sum_{n=1}^N \frac{1}{N}(a \cdot w_n + b)(1 - f_n) = \sum_{n=1}^N \frac{1}{N}(a \cdot w_n + b)(1 - f_n'), \qquad (11)$$

---

[3] The quicksort algorithm [23] is used, which averagely consumes 0.05ms (3.2×10^{-4}mWh) per frame.

then the following inequality holds

$$\sum_{n=1}^N \frac{1}{N} l \cdot w_n \cdot (1 - f_n) \leq \sum_{n=1}^N \frac{1}{N} l \cdot w_n \cdot (1 - f_n'). \qquad (12)$$

Proof: When $\mathbf{F}' = \mathbf{F}$, obviously the following equation holds exactly,

$$\sum_{n=1}^N \frac{1}{N} l \cdot w_n \cdot (1 - f_n) = \sum_{n=1}^N \frac{1}{N} l \cdot w_n \cdot (1 - f_n'). \qquad (13)$$

When $\mathbf{F}' \neq \mathbf{F}$, if $I < \sum_{n=1}^N (1 - f_n')$, then the following inequality exists,

$$\sum_{n=1}^I \widetilde{w}_n < \sum_{n=1}^N w_n \cdot (1 - f_n'), \qquad (14)$$

due to the fact that $\widetilde{w}_I$ is the I-th value of ascending sorted $\{w_n\}_{n=1}^N$. So, the following inequality holds in the case $a > 0$ and $b > 0$,

$$\sum_{n=1}^I a \cdot \widetilde{w}_n + I \cdot b < \sum_{n=1}^N a \cdot w_n \cdot (1 - f_n') + \sum_{n=1}^N b \cdot (1 - f_n'). \qquad (15)$$

Because of (10), $I = \sum_{n=1}^N (1 - f_n)$ holds, then (11) can be rewritten as

$$\sum_{n=1}^I a \cdot \widetilde{w}_n + I \cdot b = \sum_{n=1}^N a \cdot w_n \cdot (1 - f_n') + \sum_{n=1}^N b \cdot (1 - f_n'), \qquad (16)$$

which contradicts with (15). Therefore, we have $I \geq \sum_{n=1}^N (1 - f_n')$. As such, given $w_n \in [0, 1]$ and $l > 0$, we can obtain the following inequality,

$$\sum_{n=1}^I l \cdot \widetilde{w}_n = \sum_{n=1}^N l \cdot w_n \cdot (1 - f_n) \leq \sum_{n=1}^N l \cdot w_n \cdot (1 - f_n'). \qquad (17)$$

Finally, the inequality (12) can be achieved upon (17).

This completes the proof of Lemma 1.

According to Lemma 1, if and only if $w_n \leq \widetilde{w}_I$, $f_n = 0$ is the optimized solution to (9). Considering that $\sum_{n=1}^N \frac{1}{N}(a \cdot w_n + b)(1 - f_n)$ should be as close to ΔC_T as possible, the optimized solution for SQOCC approach can be expressed as

$$f_n = \begin{cases} 0, & w_n \leq \widetilde{w}_I \\ 1, & \text{otherwise}, \end{cases} \qquad (18)$$

where I satisfies

$$\frac{1}{N} \sum_{n=1}^I (a \cdot \widetilde{w}_n + b) \geq \Delta C_T > \frac{1}{N} \sum_{n=1}^{I-1} (a \cdot \widetilde{w}_n + b). \qquad (19)$$

4. EXPERIMENTAL RESULTS

In this section, experimental results are provided to validate the effectiveness of our SQOCC approach, in comparison with the latest HEVC decoding complexity reduction approach on [12]. First, the settings of our experiments are presented in Section 4.1. Then, the effectiveness of our SQOCC approach is evaluated in two aspects: the accuracy of decoding complexity control (Section 4.2) and complexity-distortion performance (Section 4.3).

4.1. Settings

We tested our approach on three 1920×1080 sequences *ParkScene*, *Kimono* and *BQTerrace* from Class B, two 832×480 sequences *PartyScene* and *RaceHorses* from Class C, and three 1280×720 conversational sequences *FourPeople*, *Johnny* and *Vidyo4* from Class E/E′. First, all the tested sequences were encoded by the HM 16.0 encoder. Here, the Random Access (RA) configuration was implemented with Group Of Picture (GOP) being 8. Two common QPs, 27 and 32, were set to encode the test videos without any rate control. Besides, the deblocking filters were enabled during the encoding. Note that all other parameters of the encoder were set by default. For decoding, HM 16.0 was also utilized with the default settings.

The experiments were all performed on a Windows PC with Inter(R) Core(TM) i7-4790K CPU and 32G RAM. In order to measure the decoding complexity, Intel® Power Gadget 3.0 was used to record the decoding power consumption in terms of mWh. In order to evaluate visual quality loss, both objective Y-PSNR difference (ΔPSNR) and subjective Y-PSNR difference (Δ_SPSNR) were measured. For fair comparison, we utilized human fixation maps from eye-tracking experiments, instead of our detected saliency maps, to weight MSE for calculating Δ_SPSNR.

4.2. Accuracy of Complexity Control

Table 3 demonstrates the accuracy of HEVC decoding complexity control of our SQOCC approach. Note that the computational complexity of saliency detection and quicksort, which takes up much less complexity than HEVC decoding, is included in our decoding complexity. Also, note that we do not compare with [12] in terms of control accuracy, since [12] is a complexity reduction approach, rather than complexity control, for HEVC decoding. We can see from Table 3 that the absolute control errors of most sequences are less than 1.00%. Besides, the Mean Absolute Errors (MAEs) are 0.35%, 1.04%, 0.50%, 0.78% at $\Delta C_T = 3.00\%, 8.00\%, 12.00\%, 15.00\%$ for QP = 27. Similar MAEs can be found for QP = 32. In general, the results here verifies that our SQOCC approach is able to control the HEVC decoding complexity with high accuracy.

4.3. Performance of quality loss

In this subsection, we evaluate ΔPSNR and Δ_SPSNR of our SQOCC approach, compared to those of [12]. The results of conversational and non-conversational sequences are shown in Figs. 4 and 5. It can be seen from these two figures that our approach significantly outperforms [12] in terms of both ΔPSNR and Δ_SPSNR. We can further see from these figures that Δ_SPSNR is less than ΔPSNR in our approach, when complexity reduction is below $\sim 10\%$. This verifies that the subjective quality can be optimized in our SQOCC approach. Once the complexity reduction becomes larger, the subjective quality may not be ensured, since much more deblocking
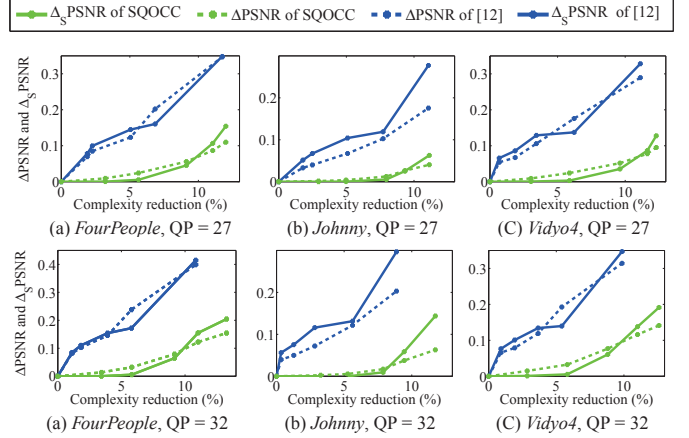
Fig. 5. ΔPSNR and Δ_SPSNR for conversational sequences.

Kimono, QP = 32, $\Delta C = 9.17\%$, SQOCC
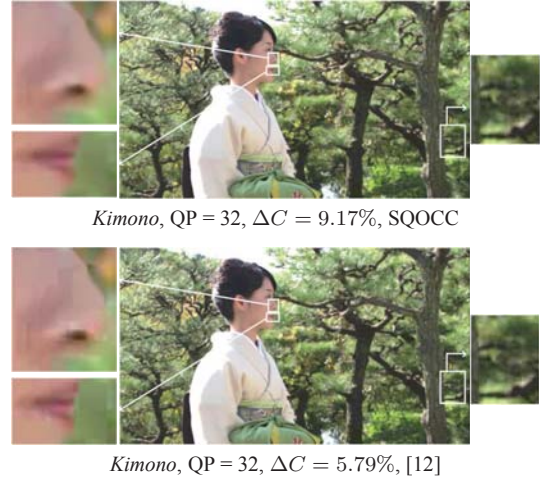
Kimono, QP = 32, $\Delta C = 5.79\%$, [12]

Fig. 6. The 36-th frame of *Kimono* decoded by our SQOCC and [12] approach.

filters in ROIs are disabled. Note that when complexity reduction reaches the maximum, Δ_SPSNR may be larger than ΔPSNR. This is because the distortion in ROIs is larger in comparison with that of non-ROIs, and Δ_SPSNR imposes the distortion of ROIs with higher weights.

Furthermore, Fig. 6 shows one selected frame of *Kimono* at QP = 32, decoded by HM 16.0 with our approach and the conventional approach [12]. Clearly, our approach produces higher visual quality in face regions than [12], even when the complexity reduction is much more in our approach. Besides, the non-ROI background is with low quality in our approach, as there exist obvious blocky effects. This also indicates that our approach is capable of optimizing subjective quality by maintaining high quality in ROIs at the expense of low quality in non-ROIs.

5. CONCLUSION

In this paper, we have proposed a novel approach, named SQOCC, for complexity control of HEVC decoding. Firstly, for subjective quality assessment, we proposed to detect video saliency, according to bit allocation in HEVC domain. Next, we developed an optimization formulation which controls HEVC decoding complexity to a target with minimal

Table 3. Complexity control performance of SQOCC approach

QP	ΔC_T (%)	Non-conversational sequences					Conversational sequences			MAE (%)
		ParkScene	*Kimono*	*BQTerrace*	*PartyScene*	*RaceHorses*	*FourPeople*	*Johnny*	*Vidyo4*	
27	3.00	2.99	2.96	3.36	2.45	4.55	3.21	2.94	3.04	0.35
	8.00	9.43	8.95	9.77	6.88	8.23	9.12	7.94	9.61	1.04
	12.00	12.75	12.06	12.11	10.30	11.79	12.00	11.13	12.29	0.50
	15.00	14.33	14.23	16.18	15.81	15.92	-	-	-	0.87
32	3.00	2.81	3.27	4.55	2.02	3.10	3.43	3.26	2.85	0.49
	8.00	8.17	8.39	9.64	7.59	7.81	9.17	7.88	8.80	0.82
	12.00	12.51	11.95	11.62	9.42	11.53	11.74	13.26	12.59	0.76
	15.00	14.89	14.49	16.14	14.28	15.30	-	-	-	0.56
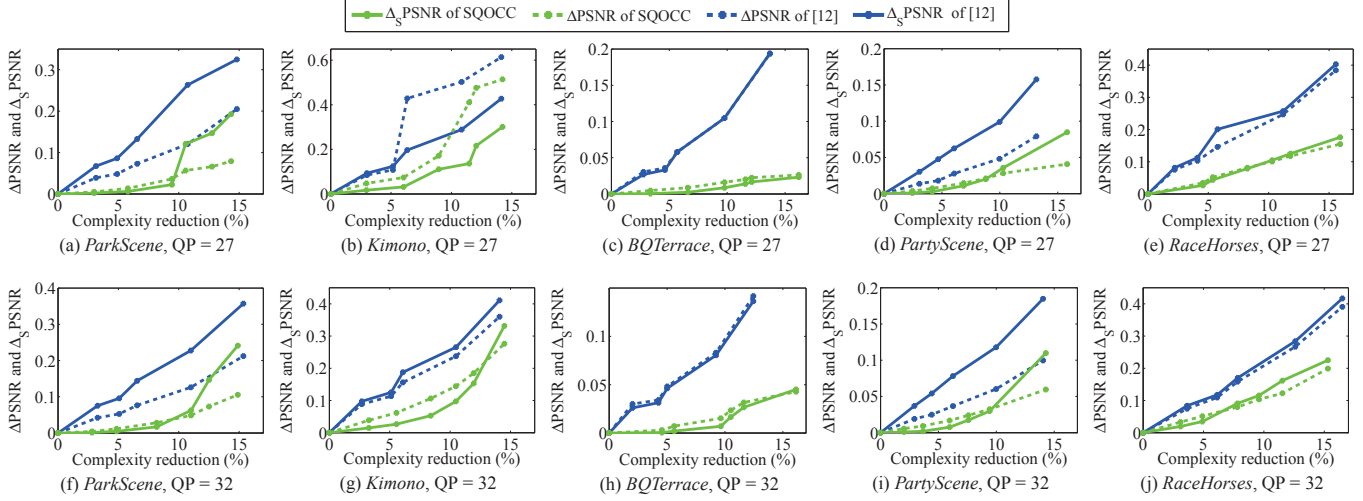
Fig. 4. ΔPSNR and Δ_SPSNR for non-conversational sequences.

subjective quality loss, via disabling the deblocking filters of some specific LCUs. Then, we investigated the relationship between saliency and subjective quality loss, and the relationship between saliency and deblocking complexity. Finally, we provided a solution to our optimization formulation, achieving SQOCC for HEVC decoding. The experimental results validate the effectiveness of our SQOCC approach.

6. REFERENCES

[1] Gary J Sullivan, J-R Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE TCSVT*, pp. 1649–1668, 2012.

[2] Frank Bossen, Benjamin Bross, Karsten Suhring, and Damian Flynn, "HEVC complexity and implementation analysis," *IEEE TCSVT*, pp. 1685–1696, 2012.

[3] Mauricio Alvarez-Mesa, Chi Ching Chi, Ben Juurlink, Valeri George, and Thomas Schierl, "Parallel video decoding in the emerging HEVC standard," in *ICASSP*, 2012.

[4] Chi Ching Chi, Mauricio Alvarez-Mesa, Ben Juurlink, Gordon Clare, Félix Henry, Stéphane Pateux, and Thomas Schierl, "Parallel scalability and efficiency of HEVC parallelization approaches," *IEEE TCSVT*, pp. 1827–1838, 2012.

[5] Leju Yan, Yizhou Duan, Jun Sun, and Zongming Guo, "Implementation of HEVC decoder on x86 processors with SIMD optimization," in *VCIP*, 2012.

[6] Chi Ching Chi, Mauricio Alvarez-Mesa, Benjamin Bross, Ben Juurlink, and Thomas Schierl, "SIMD acceleration for HEVC decoding," *IEEE TCSVT*, 2014.

[7] Diego F de Souza, Aleksandar Ilic, Nuno Roma, and Leonel Sousa, "Towards GPU HEVC intra decoding: Seizing fine-grain parallelism," in *ICME*, 2015.

[8] Erwan Nogues, Romain Berrada, Maxime Pelcat, Daniel Menard, and Erwan Raffin, "A DVFS based HEVC decoder for energy-efficient software implementation on embedded processors," in *ICME*, 2015.

[9] Yang Liu, Zheng Guo Li, and Yeng Chai Soh, "Region-of-interest based resource allocation for conversational video communication of H.264/AVC," *IEEE TCSVT*, pp. 134–139, 2008.

[10] Matteo Naccari, Catarina Brites, João Ascenso, and Fernando Pereira, "Low complexity deblocking filter perceptual optimization for the HEVC codec," in *ICIP*, 2011.

[11] Christian Feldmann, Fabian Jager, and Mathias Wien, "Decoder complexity reduction for the scalable extension of HEVC," in *ICIP*, 2014.

[12] Erwan Nogues, Simon Holmbacka, Maxime Pelcat, Daniel Menard, and Johan Lilius, "Power-aware HEVC decoding with tunable image quality," in *SiPS*, 2014.

[13] Erwan Nogues, Erwan Raffin, Maxime Pelcat, and Daniel Menard, "A modified HEVC decoder for low power decoding," in *ACM ICCF*, 2015.

[14] Mohsen Jamali Langroodi, Joseph Peters, and Shervin Shirmohammadi, "Decoder-complexity-aware encoding of motion compensation for multiple heterogeneous receivers," *ACM TOMM*, p. 46, 2015.

[15] Ethel Matin, "Saccadic suppression: a review and an analysis," *Psychological bulletin*, pp. 899–917, 1974.

[16] Ali Borji and Laurent Itti, "State-of-the-art in visual attention modeling," *IEEE TPAMI*, pp. 185–207, 2013.

[17] Sayed Hossein Khatoonabadi, Nuno Vasconcelos, Ivan V Bajic, and Yufeng Shan, "How many bits does it take for a stimulus to be salient?," *NIPS*, 2006.

[18] J-R Ohm, Gary J Sullivan, Heiko Schwarz, Thiow Keng Tan, and Thomas Wiegand, "Comparison of the coding efficiency of video coding standards including high efficiency video coding (HEVC)," *IEEE TCSVT*, pp. 1669–1684, 2012.

[19] Chenlei Guo and Liming Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE TIP*, pp. 185–198, 2010.

[20] Dmitry Rudoy, Dan B Goldman, Eli Shechtman, and Lihi Zelnik-Manor, "Learning video saliency from human gaze using candidate selection," in *CVPR*, 2013.

[21] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE TPAMI*, pp. 1254–1259, 1998.

[22] Zhicheng Li, Shiyin Qin, and Laurent Itti, "Visual attention guided bit allocation in video compression," *IVC*, vol. 29, no. 1, pp. 1–14, 2011.

[23] Charles AR Hoare, "Quicksort," *The Computer Journal*, pp. 10–16, 1962.