2016 Data Compression Conference

Optimizing Subjective Quality in HEVC-MSP: An Approximate Closed-form Image Compression Approach

Shengxi Li[1], Mai Xu[1 2 *], Yun Ren[1], Chengzhang Ma[1], and Zulin Wang[1 2]

[1]Beihang University
37 Xueyuan Road
Beijing, 100191, China
MaiXu@buaa.edu.cn

[2]Collaborative Innovation Center of Geospatial Technology
129 Luoyu Road
Wuhan, 430079, China
wzulin@buaa.edu.cn

Abstract

HEVC, as the latest video coding standard, achieves top performance on image compression. On the basis of this, we propose a novel approach to optimize subjective quality for HEVC-based image compression. Specifically, a bit allocation formulation is established to optimize subjective quality with constraint on bit-rates. Then, we propose a recursive Taylor expansion method to quickly solve such a formulation with an approximate closed-form solution. The experimental results show the superior performance of our approach, with ∼40% BD-rate saving over the state-of-the-art HEVC-MSP for face image compression.

1 Introduction

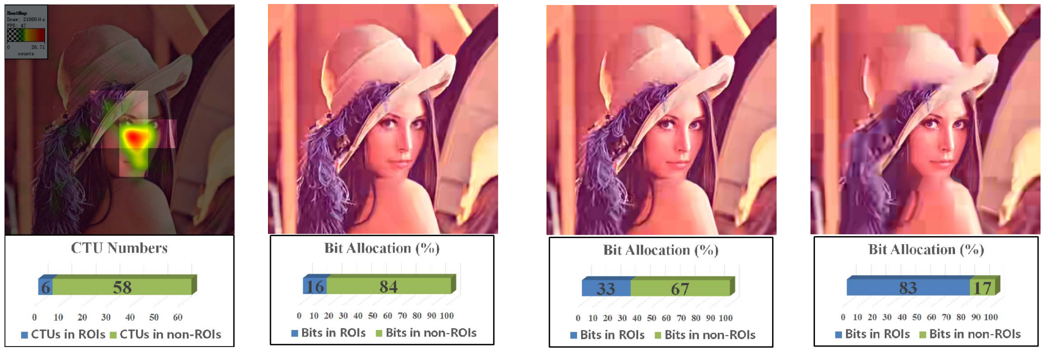
Nowadays, multimedia applications, such as Facebook and Twitter, are becoming the integral component for the daily life of millions, leading to the explosion of big data. Among them, images account for an important part, thus posing great challenges to the limited communication bandwidth. A set of image compression standards have been proposed to condense the image data, e.g., JPEG 2000, JPEG XR, and WebP. Benefitting from most recent success of high efficiency video coding (HEVC) [1], HEVC main still picture (HEVC-MSP) profile [2] achieves performs the best among all state-of-the-art standards on image compression [2]. However, those existing standards, including HEVC-MSP, mainly focus on removing statistical redundancy with various techniques [3]. Further reducing the statistical redundancy may help to improve coding efficiency, but at the cost of extremely heavy computational complexity.

Thanks to the human visual system (HVS), there exists perceptual redundancy in images that can be further exploited [4]. Many ongoing approaches on mimicking the HVS have cascaded some light on perceptual image compression. The features developed from the HVS can be mainly classified into two categories [3]: visual sensitivity and visual attention. Correspondingly, the existing approaches are mainly related to: **I**, bit reduction while maintaining a desired subjective quality (visual sensitivity); **II**, quality improvement with the constraint on bit-rates (visual attention).

For **I**, one commonly used HVS feature is the just noticeable difference (JND). Many approaches [5] [6] [7] [8] incorporate JND to save bits while the compressed images maintain almost unchanged or desired subjective quality. For instance, Liu *et al.* [8] adopted JND and a spatial and spectral quantization error to estimate perceptual distortion. Then, by iterating to reach the desired distortion, minimum bits can be achieved for JPEG 2000. However, JND-based approaches target at saving

∗ The corresponding author of this paper is Mai Xu.

1068-0314/16 $31.00 © 2016 IEEE
DOI 10.1109/DCC.2016.10



(a) Heat map (b) No emphasis (c) Well balanced (d) More emphasis

Figure 1: Different bit allocation emphasis on ROIs of Lena image. Note that (a) is the heat map of eye fixations; (b), (c) and (d) are compressed by HEVC-MSP and the bit-rates remain the same at 0.1 bpp. The DMOS scores (to be discussed in Section 4) for (b), (c), and (d) are 63.9, 57.5, and 70.3, respectively.

some bit-rates at a given subjective quality, and they can hardly satisfy the bit-rate constraint. For **II**, catering for visual attention of the HVS, most approaches [9] [10] [11] [12] are developed to arrange relatively more bits to the region-of-interest (ROI) [3] for better subjective quality. For example, in the work of [12], perceptual image compression is achieved by maintaining the DWT coefficients in ROIs, while reducing some coefficients in non-ROIs.

However, the extremely low quality in non-ROIs may also affect the whole image quality, as illustrated in Figure 1. Thus, how many bits should "move" from non-ROIs to ROIs is crucial for improving subjective quality. One effective way is to optimize subjective quality, which has been considered in several works [13] [14]. However, it is intractable [15] to establish a closed-form relationship between bit-rates and perceptual distortion metrics, leading to sub-optimal results during bit allocation. Since the state-of-the-art HEVC-MSP includes many delicate components, it is even more intractable to establish this relationship. To our best knowledge, there exists few bit allocation work to optimize subjective quality with the constraint on bits, especially for HEVC-MSP.

Therefore, we propose in this paper a novel bit allocation approach to optimize subjective quality for HEVC-based image compression. As pointed out by the latest work [16], the information content weight peak signal-to-noise ratio (IW-PSNR), which simply combines pixel-wise saliency with mean squared error (MSE), is effective to maintain high correlation with subjective quality. We thus apply IW-PSNR in this paper to model subjective quality. Then, we propose a bit allocation formulation to optimize IW-PSNR at a given bit-rate. Unfortunately, it is intractable to obtain a closed-form solution to this formulation. Then, the recursive Taylor expansion (RTE) is proposed to acquire the approximate closed-form solution. Additionally, to deal with the mismatch between target and actual bits, we develop an optimal bit reallocation process to accurately control bit-rate while maintaining optimization. As verified in our analysis, little complexity is introduced in our approach.

2 Optimizing subjective quality

In this section, we mainly focus on optimizing subjective quality. To this end, Section 2.1 first transplants the R-λ rate control (RC) approach [17] into HEVC-MSP. Based on this, Section 2.2 proposes an optimization formulation and Section 2.3 solves this formulation with an approximate closed-form solution. Section 2.4 develops an optimal bit re-allocation method to solve the issue of mismatch between the target and actual bits.

2.1 Rate control implementation on HEVC-MSP

The latest R-λ approach is proposed in [17] for RC in HEVC. Since we concentrate on applying RC to image compression, the CTU level RC in one video frame is mainly discussed here. Specifically, for HEVC, it has been verified that the Hyperbolic model can better fit rate-distortion (R-D) relationship [17]. Based on this, an R-λ model is utilized for bit allocation in the latest RC approach, where λ is the slope of R-D relationship [18]. Assuming that d_i, r_i and λ_i represent the distortion, bits and R-D slope for the i-th CTU, the R-D relationship and R-λ model are formulated as follows,

$$d_i = c_i {r_i}^{-k_i}, \qquad \lambda_i = -\frac{\partial d_i}{\partial r_i} = c_i k_i \cdot {r_i}^{-k_i-1}, \tag{1}$$

where c_i and k_i are the constants reflecting the content of the i-th CTU. In the R-λ approach [17], r_i is first allocated according to the predicted mean absolute difference (MAD), and then its corresponding λ_i is obtained with (1). By adopting a fitting relationship between λ_i and QP_i, QPs of all CTUs within the frame can be yielded, such that RC is achieved in HEVC. For more details, refer to [17].

However, c_i and k_i cannot be obtained when encoding the current CTU for HEVC-MSP. Thus, it is hard to directly apply R-λ RC to HEVC-MSP. To predict the image content, Karczewicz *et al.* [19] proposed to adopt sum of absolute transformed difference (STAD) with constant c_i and k_i for R-λ RC of HEVC-MSP. Although SATD can reflect the texture complexity, it is not as good as c_i and k_i on representing image content. Therefore, this results in inaccurate R-D relationship and degradation on coding efficiency.

To avoid above issues, we adopt the pre-processing process in calculating c_i and k_i. After pre-compressing, the pre-encoded distortion, bits and λ can be obtained for the i-th CTU, which are denoted as $\bar{d}_i$, $\bar{r}_i$ and $\bar{\lambda}_i$. Then, the RC related parameters, c_i and k_i, can be estimated with (1) before encoding the i-th CTU:

$$c_i = \bar{d}_i / \left(\bar{r}_i^{-\bar{\lambda}_i \cdot \bar{r}_i / \bar{d}_i} \right), \qquad k_i = \frac{\bar{\lambda}_i \cdot \bar{r}_i}{\bar{d}_i}. \tag{2}$$

This way, with c_i and k_i, the RC of the R-λ approach [17] can be efficiently implemented in HEVC-MSP.

Here, a fast pre-compressing process is developed in our approach, which set the maximum coding unit (CU) depth to 0 for all pre-compressing CTUs. We have verified that our developed pre-compressing process increases the computational complexity by 10% burden, which is a bit more than 6% of the SATD-based method [19].

2.2 Optimization formulation on subjective quality

The main target of this paper is to optimize the subjective quality for HEVC-based image compression. Here, the subjective quality is approximated by IW-PSNR [16], as [16] has shown that IW-PSNR is highly correlated with subjective quality. In general, IW-PSNR weights the distortion of each pixel by their corresponding saliency values. Here, we denote d_i and w_i as the sum of pixel-wise distortion (MSE) and saliency values within the i-th CTU. Then, on the basis of d_i and w_i, the maximization on IW-PSNR at a given target bit-rate R can be formulated by

$$\min\left(\frac{\Sigma_{i=1}^{M} w_i d_i}{\Sigma_{i=1}^{M} w_i}\right) \quad \text{s.t.} \;\; \Sigma_{i=1}^{M} r_i = R. \tag{3}$$

By using Lagrange multiplier λ and R-D cost J [18], (3) can be turned to minimize J. Then, by setting its derivative to zero, the minimum J can be solved:

$$\frac{\partial J}{\partial r_i} = \frac{\partial\left(\Sigma_{i=1}^{M} w_i d_i / \Sigma_{i=1}^{M} w_i + \lambda(\Sigma_{i=1}^{M} r_i)\right)}{\partial r_i} = \frac{w_i}{\Sigma_{i=1}^{M} w_i} \cdot \frac{\partial d_i}{\partial r_i} + \lambda = 0. \tag{4}$$

Combining R-D relationship in (1), (4) is turned to

$$r_i = \left(\frac{\lambda \cdot \Sigma_{i=1}^{M} w_i}{c_i k_i w_i}\right)^{-\frac{1}{k_i+1}} = \left(\frac{\widetilde{w}_i a_i}{\lambda}\right)^{b_i}, \tag{5}$$

where $a_i = c_i k_i$ and $b_i = \frac{1}{k_i+1}$, reflecting the image content of each CTU. Moreover, $\widetilde{w}_i = w_i/(\Sigma_{i=1}^{M} w_i)$ represents the visual importance for each CTU. Note that with our pre-compressing process, c_i and k_i can be obtained in advance. Thus, a_i and b_i are able to be calculated before encoding the image. Then, the minimum J can be achieved once λ is known in (5).

For calculating λ, we can use the constraint on bit-rates, formulated as $\sum_{i=1}^{M} r_i = R$. In other words, we need to find the "proper" λ by,

$$\sum_{i=1}^{M} r_i = \sum_{i=1}^{M} \left(\frac{\widetilde{w}_i a_i}{\lambda}\right)^{b_i} = R. \tag{6}$$

After solving (6) to find the "proper" λ, the target bits satisfies the bit-rate constraint, meanwhile enjoying minimal J and maximal IW-PSNR for the compressed image.

Unfortunately, since a_i and b_i vary across different CTUs, (6) cannot be solved by a closed-form solution. Therefore, in the next section, the RTE method is proposed to provide an approximate closed-form solution towards (6).

2.3 RTE method for solving optimization formulation

For solving (6), we assume that $\widetilde{r}_i(\widetilde{\lambda})^{b_i} = (\widetilde{w}_i a_i)^{b_i}$, where $\widetilde{r}_i$ and $\widetilde{\lambda}$ are the estimated r_i and λ, respectively. Then, (6) can be rewritten as,

$$\sum_{i=1}^{M} r_i = \sum_{i=1}^{M} \left(\frac{\widetilde{w}_i a_i}{\lambda}\right)^{b_i} = \sum_{i=1}^{M} \widetilde{r}_i \left(\frac{\widetilde{\lambda}}{\lambda}\right)^{b_i} = R. \tag{7}$$

From (7), we can see that once $\widetilde{\lambda} \to \lambda$, there exists $\widetilde{r}_i \to r_i$. As such, the optimization formulation of (6) can be solved for the bit allocation towards optimal subjective quality. However, we do not know $\widetilde{\lambda}$ at the beginning. Meanwhile, λ of (7) is unknown, as it is intractable to find the closed-form solution to (6). In other words, there exists a chicken-and-egg dilemma between $\widetilde{\lambda}$ and λ. To solve such a dilemma, a possible $\widetilde{\lambda}$ is initially set[1]. Then, the RTE method is proposed to iteratively update $\widetilde{\lambda}$ for $\widetilde{\lambda} \to \lambda$.

Specifically, we preliminarily apply Taylor Expansion on $(\frac{\widetilde{\lambda}}{\lambda})^{b_i}$ of (7), and then discard the biquadratic and higher-order terms. The process can be formulated as

[1] The picture λ [17] is set as the initial value of $\widetilde{\lambda}$ in our RTE method.

$$R=\sum_{i=1}^{M}\widetilde{r}_i(\frac{\widetilde{\lambda}}{\lambda})^{b_i}\approx\sum_{i=1}^{M}\widetilde{r}_i\left(1+\frac{\ln(\frac{\widetilde{\lambda}}{\lambda})}{1!}b_i+\frac{(\ln\frac{\widetilde{\lambda}}{\lambda})^2}{2!}b_i^2+\frac{(\ln\frac{\widetilde{\lambda}}{\lambda})^3}{3!}b_i^3\right) \tag{8}$$

$$=\underbrace{-\sum_{i=1}^{M}\widetilde{r}_i(\frac{b_i^3}{6})}_{A}\ln^3\lambda+\underbrace{\sum_{i=1}^{M}\widetilde{r}_i(\frac{b_i^2}{2}+\frac{b_i^3}{2}\ln\widetilde{\lambda})}_{B}\ln^2\lambda\underbrace{-\sum_{i=1}^{M}\widetilde{r}_i(b_i^2\ln\widetilde{\lambda}+b_i+\frac{b_i^3}{2}\ln^2\widetilde{\lambda})}_{C}\ln\lambda+\underbrace{\sum_{i=1}^{M}\widetilde{r}_i(1+b_i\ln\widetilde{\lambda}+\frac{b_i^2}{2}\ln^2\widetilde{\lambda}+\frac{b_i^3}{6}\ln^3\widetilde{\lambda})}_{D}.$$

Then, (7) can be approximated to be the cubic equation with variable $\ln\lambda$ in (8). Applying Shengjin formula [20], this cubic equation is worked out to obtain the approximated solution λ (denoted by $\widehat{\lambda}$) as:

$$\widehat{\lambda}=e^{\frac{-B-(\sqrt[3]{Y_1}+\sqrt[3]{Y_2})}{3A}},\quad Y_{1,2}=BE+3A(\frac{-F\pm\sqrt{F^2-4EG}}{2}), \tag{9}$$

where $E=B^2-3AC$, $F=BC-9A(D-R)$, and $G=C^2-3B(D-R)$. Note that as $\Delta=F^2-4EG>0$ in practical encoding, there exists only one real root [20] for (9). Thus, $\widehat{\lambda}$ value is unique. However, due to the truncation of higher-order terms in Taylor expansion, $\widehat{\lambda}$ estimated by (9) is not the accurate solution to (7). Fortunately, as proved in Lemma 1[2], $\widehat{\lambda}$ is more accurate for estimating λ than $\widetilde{\lambda}$ when $\widetilde{\lambda}<\lambda$.

Lemma 1 *Consider $\lambda>\widetilde{\lambda}>0$, $b_i>0$, and $R>0$, for (7). When the solution of λ to (7) is $\widehat{\lambda}$, the following inequality holds for $\widehat{\lambda}$,*

$$|\widehat{\lambda}-\lambda|<|\widetilde{\lambda}-\lambda|. \tag{10}$$

Proof *See our website*[3].

As seen from Lemma 1, although both $\widetilde{\lambda}$ and $\widehat{\lambda}$ may be inaccurate, $\widehat{\lambda}$, obtained through (7) - (9), is more close to λ than $\widetilde{\lambda}$. This important characteristic contributes a lot to our estimation on $\widetilde{\lambda}$ when λ of (7) is unknown. In fact, we can iterate the Taylor expansion via utilizing $\widehat{\lambda}$ as the estimation of $\widetilde{\lambda}$ to the next iteration, which is the core of RTE method. In addition, if $\widetilde{\lambda}>\lambda>0$ at the first iteration, its output $\widehat{\lambda}$ is smaller than λ, as pointed out by Lemma 2. For the subsequent iterations of the RTE method, $0<\widetilde{\lambda}<\lambda$ can be achieved, since the value of $\widetilde{\lambda}$ has been replaced by that of $\widehat{\lambda}$. This way, the approximate closed-form solution λ can be obtained, by iteratively estimating $\widetilde{\lambda}$.

Lemma 2 *Consider $\widetilde{\lambda}>0$, $\lambda>0$, $b_i>0$, $\lambda\neq\widetilde{\lambda}$, and $R>0$ for (7). If $\widehat{\lambda}$ is the solution of λ to (7), then the following holds*

$$\widehat{\lambda}<\lambda. \tag{11}$$

proof *See our website.*

Our RTE method is summarized in Table 1. For each iteration, the convergence criterion is set as the extremely low approximation error, $E_a<10^{-10}$, where $E_a=|\Sigma_{i=1}^{M}\widetilde{r}_i-R|/R$. As analyzed in Section 3, generally with no more than three iterations, the RTE method is able to reduce the difference between $\widetilde{\lambda}$ and λ to an extremely small value, meeting the convergence criterion. Thus, $\widetilde{\lambda}$ can be output as the approximate closed-form solution to (7) (as well as (6)). Finally, we replace λ by $\widetilde{\lambda}$ in (5) to allocate the target bits to each CTU, such that IW-PSNR can be maximized.

[2] It is obvious that $0<b_i=\frac{1}{k_i+1}<1$ and $R>0$ in HEVC encoding.

[3] Our website is http://www.ee.buaa.edu.cn/xumfiles/dcc2016.htm

Table 1: The RTE method for solving (7)

- **Input:** Image content information (a_i, b_i, w_i) for each encoding CTUs and target bits R for the image.
- **Output:** reasonable bit allocation $\widetilde{r}_i$ for each CTU on optimizing subjective quality.
- Initialize pre-estimated $\widetilde{\lambda}$ to be the picture λ.
- **While** $\widetilde{\lambda}$ does not meet the convergence criterion
  1. Calculate A, B, C, and D of (8) with $\widetilde{\lambda}$.
  2. Obtain $\widehat{\lambda}$ estimated by (9).
  3. Update $\widetilde{\lambda}$ with the obtained $\widehat{\lambda}$.

  End
- Save the final $\widetilde{\lambda}$.
- Apply it to bit allocation $\widetilde{r}_i$ with (5).
- **Return** $\widetilde{r}_i$ for each CTU.

2.4 Bit re-allocation for maintaining optimization

As we have discussed in Section 2.3, bits are reasonably allocated in our method to optimize subjective quality. However, in practical encoding, there may exist slight difference between target and actual bits for each CTU. This difference may degrade the control accuracy as well as optimization on subjective quality. To address such an issue, we develop a bit re-allocation process to accurately control bits, meanwhile maintaining the optimization on subjective quality.

Specifically, for compensating the bit error after encoding the i-th CTU, the target bits for the incoming K CTUs (denoted as $T_{i+1,i+K}$) are updated by

$$T_{i+1,i+K} = \sum_{j=i+1}^{j=i+K} \widetilde{r}_j + \underbrace{\left(\widehat{T} - \sum_{j=i+1}^{j=M} \widetilde{r}_j \right)}_{\text{bit error}}. \tag{12}$$

In (12), $\widehat{T}$ is the left bits for encoding remaining CTUs. Recall that M means the total number of CTUs, and $\widetilde{r}_j$ represents the target bits for the j-th CTU by our RTE method. Obviously, as seen from (12), the mismatch between target and actual bits is compensated during encoding the next K CTUs. Here, the RTE method of Section 2.3 is applied to re-allocate $T_{i+1,i+K}$ to the next K CTUs. Note that we follow [17] to set $K = 4$, which means that bits are re-assigned in the next four CTUs. Moreover, it needs to be pointed out that due to the fast convergence speed of our RTE method, the complexity increases little for the bit re-allocation process.

Finally, we summarize our approach in Figure 2. Specifically, we first transplant R-λ RC into HEVC-MSP with a simplified pre-compressing process, and the saliency values are detected for the encoding image. Based on this, the RTE method is used to design bits to each CTU, such that subjective quality can be optimized at a given bit-rate. Next, the QP value of each CTU can be estimated upon the R-λ model and QP fitting. Note that the bits need to be re-allocated in the following CTUs, to bridge the gap between target and actual bits.

3 Computational complexity analysis

This section discusses the computational complexity of our approach. Figure 3 shows E_a versus RTE iterations when applying our approach to image compression in HM 16.0 platform. From this figure we can see that with at most three iterations, E_a reaches below to 10^{-10}, implying the fast convergence speed of our RTE method.

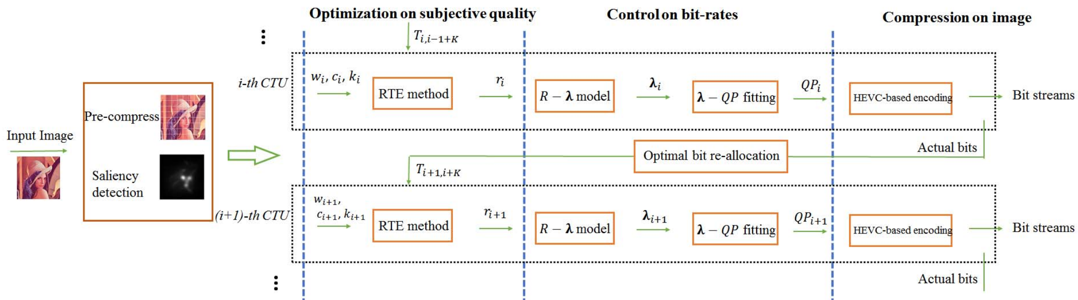
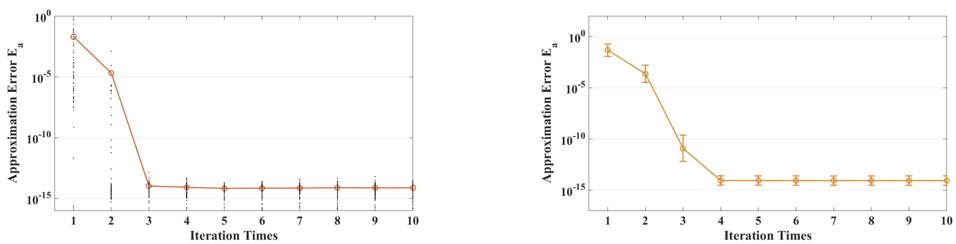


Figure 2: The procedure of our approach on optimizing subjective quality.



(a) Lena 512×512 @ 512bpp (b) All images of our test set

Figure 3: E_a versus iteration times of the RTE method. Note that for (a), the black dots represent $\Delta\lambda$ for each CTU in Lena image. For (b), all 10 images (from our test set of Section 4) were used to calculate the approximation error and the corresponding standard deviation along with the increased iterations.

We further exploit the computational time for each iteration of the RTE method. As seen from Table 1, the computational time for each iteration is independent of image content. Thus, one image was randomly chosen and compressed by our approach. The averaged time of one iteration of our RTE method was then recorded. The computer used for the test is with Intel Core i7-4770 CPU at 3.4 GHz and 16 GB RAM. Through the test, one iteration of our RTE method only consumes around 0.0015 ms for each CTU. Since it takes at most three iterations to acquire the approximate closed-form, the computational time for each solution is less than 0.005 ms in our approach.

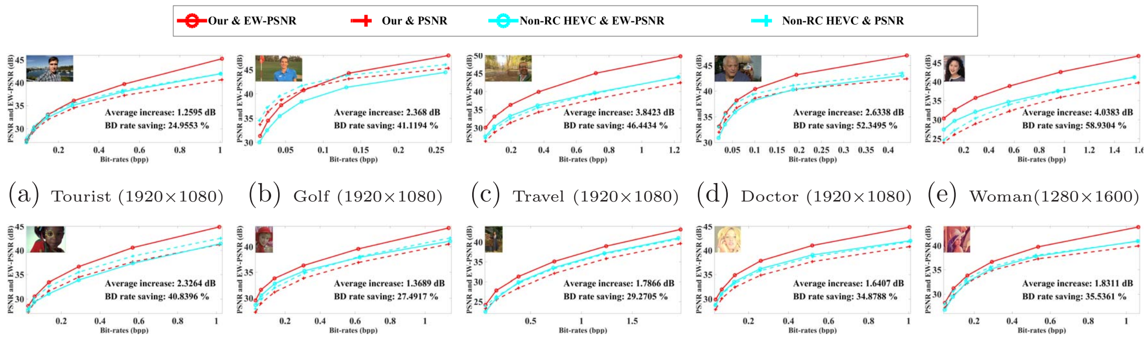
Our approach consists of two parts: bit allocation and re-allocation with the RTE method. For bit allocation, three iterations are enough for encoding one image, thus consuming at most 0.005 ms. For bit re-allocation, the computational time depends on CTU number of the image, and each CTU requires at most three iterations to obtain the re-allocated bits. For a 1080p image, the computational time is around 2.5 ms as it includes 510 CTUs. This implies the negligible computational complexity burden of our approach.

4 Experimental results

Experimental results are reported in this section to evaluate the performance of our approach. As face images take a large part of images in daily life and human face is consistently agreed to draw much attention, the face images are used in our experiment to evaluate the performance of our approach. Here, we follow our most recent work [21] to detect saliency and obtain w_i for our optimization formulation (3).

4.1 Test set and parameter settings

We set up a test set which includes 10 face images at different resolutions and scenarios (available at our website). For this test set, two standard test images (*Lena* and



(a) Tourist (1920×1080) (b) Golf (1920×1080) (c) Travel (1920×1080) (d) Doctor (1920×1080) (e) Woman(1280×1600)

(f) Kodim15 (768×512) (g) Kodim04 (512×768) (h) Kodim18 (512×768) (i) Tiffany (512×512) (j) Lena (512×512)

Figure 4: EW-PSNR and PSNR versus bit-rates for our approach and conventional non-RC HEVC-MSP.

Tiffany) were selected and all face images from JPEG XR and Kodak were chosen (*Woman*, *Kodim04*, *Kodim15*, and *Kodim18*). Besides, four images (*Tourist*, *Golf*, *Travel*, and *Doctor*) were chosen from our eye-tracking database [21]. For more details of these images, refer to Figure 4.

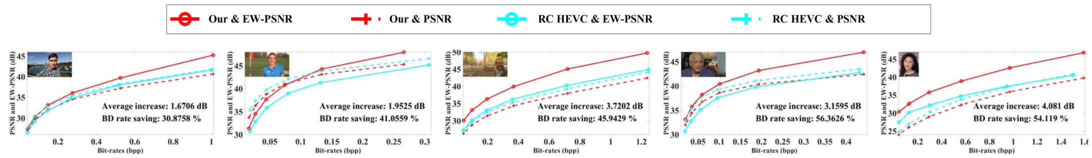
The non-RC HEVC-MSP [2], with the default MSP configuration profile on HM 16.0 platform, was utilized for comparison. The RC HEVC-MSP, mainly based on [19], was also compared. Note that our approach and the RC HEVC-MSP have added RC to specify the bit-rates, and the other parameters in the configuration profile were set by default, the same as those of the non-RC HEVC-MSP. To obtain the target bit-rates, we encoded each image with the non-RC HEVC-MSP at fixed QPs, of which values are 47, 42, 37, 32, 27, and 22. As such, high ranges of compressed image quality can be ensured. Then, the target bit-rates of our approach and the RC HEVC-MSP were set to be the actual bits obtained by the non-RC HEVC-MSP at above QP values.

4.2 Evaluation

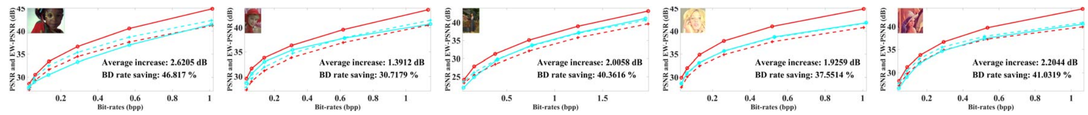
Now, we assess on visual quality of our approach, the non-RC and RC HEVC-MSP. Since eye fixations acquired by the eye-tracking experiment[4] can well reflect visual attention, it is more reasonable to use saliency generated by eye fixations, instead of saliency by our method [22], for visual quality assessment (VQA). Accordingly, we compare our approach with the non-RC and RC HEVC-MSP in the terms of eye-tracking weight PSNR (EWPSNR) [22], whose effectiveness is verified in VQA.

Then, Figures 4 and 5 plot the EWPSNR and PSNR versus bit-rates for each image. From this figure we can see that although there is slight degradation on PSNR, our approach is able to significantly improve EWPSNR, due to the proper emphasis on ROIs. Specifically, our approach enjoys averaged 2.46 dB improvement over the non-RC HEVC-MSP and even more improvement over the RC HEVC-MSP. Moreover, the Bjontegaard distortion-rate (BD-rate) saving for EW-PSNR in our approach is 38.54% over the non-RC HEVC-MSP, and 42.48% over the RC HEVC-MSP. This verifies that our approach can significantly improve subjective quality in terms of EW-PSNR, compared with the non-RC and RC HEVC-MSP.

[4]We have conducted the eye-tracking experiment on images in our test set, with test condition the same as [21]. The results of our eye-tracking experiment can also be available at our website, http://www.ee.buaa.edu.cn/xumfiles/dcc2016.htm



(a) Tourist (1920×1080) (b) Golf (1920×1080) (c) Travel (1920×1080) (d) Doctor (1920×1080) (e) Woman(1280×1600)



(f) Kodim15 (768×512) (g) Kodim04 (512×768) (h) Kodim18 (512×768) (i) Tiffany (512×512) (j) Lena (512×512)

Figure 5: EW-PSNR and PSNR versus bit-rates for our approach and conventional RC HEVC-MSP.

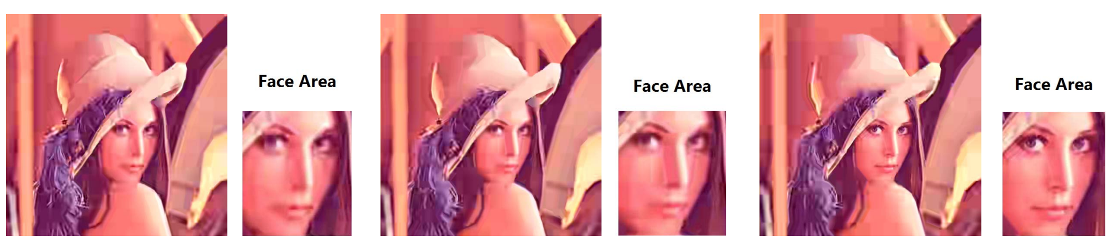
Table 2: DMOS results for our approach and conventional non-RC HEVC-MSP.

		Tourist	Golf	Travel	Doctor	Woman	Kodim15	Kodim04	Kodim18	Tiffany	Lena
QP=47	Bits (bpp)	0.04	0.02	0.04	0.02	0.04	0.03	0.03	0.05	0.03	0.05
	Our	**57.2**	**58.0**	**56.9**	**56.5**	**61.4**	**64.5**	**68.9**	**55.0**	**59.2**	**57.5**
	Non-RC	74.3	69.6	69.1	63.9	78.4	70.1	73.9	66.3	67.6	63.9
QP=42	Bits (bpp)	0.08	0.03	0.10	0.03	0.13	0.06	0.06	0.16	0.06	0.09
	Our	**45.0**	**50.0**	**42.7**	**47.8**	**43.9**	**50.7**	**53.6**	**43.1**	**43.1**	**47.9**
	Non-RC	58.5	56.3	53.7	52.1	61.3	61.2	61.9	56.9	54.1	55.5
QP=32	Bits (bpp)	0.27	0.08	0.36	0.10	0.56	0.29	0.31	0.76	0.26	0.28
	Our	**28.1**	**35.2**	**26.1**	**34.1**	**28.9**	**30.0**	**30.0**	**20.8**	**27.1**	**36.9**
	Non-RC	36.4	42.0	34.0	42.3	36.0	38.7	38.8	28.5	30.2	44.0

Furthermore, we compare our approach with the non-RC HEVC-MSP using difference mean opinion scores (DMOS). Note that the RC HEVC-MSP is not evaluated in our test, as it has the similar visual quality to non-RC HEVC-MSP. The results are tabulated in Table 2. Note that the smaller values of DMOS indicate the better subjective quality. Thus, we can see that our approach has much better subjective quality than the non-RC HEVC-MSP at all bit-rates. Moreover, for all images, the DMOS values of our approach at QP $= 47$ are almost equal to those of the non-RC HEVC-MSP at QP $= 42$. This means that nearly half reduction of bit-rates is achieved in our approach. This is also in accordance with ∼40% BD rate saving of our approach discussed above. We further show in Figure 6 *Lena* image of our results. Obviously, there exists evident visual quality improvement of our approach over the conventional non-RC and RC HEVC-MSP. We have offered all compressed images in our website.

5 Conclusion

In this paper, we have proposed a novel HEVC-based image compression approach, which optimizes the subjective quality in the latest HEVC-MSP platform. Benefitting from the state-of-the-art saliency detection, we develop a formulation to optimize subjective quality, which maintains properly high quality at attention-attracting re-



(a) Non-RC HEVC-MSP (b) RC HEVC-MSP (c) Our

Figure 6: Subjective quality of *Lena* image at 0.05 bpp (QP $= 47$) for three approaches.

gions. Then, the RTE method is proposed to solve such a formulation, followed by bit allocation and re-allocation process. As a result, the subjective quality can be significantly improved over the state-of-the-art HEVC-MSP for image compression. Our experimental results verify such a significant improvement on face image compression. Our approach, of course, is not limited to face image compression, as other saliency detection methods for generic images can be easily embedded into our approach.

6 Acknowledgement

This work was supported by the NSFC projects under Grants 61573037, 61202139, and 61471022, and the China 973 program under Grant 2013CB329006. It was also supported by the fundamental research funds for Beihang University.

References

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE TCSVT*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] T. Nguyen and D. Marpe, "Objective performance evaluation of the HEVC Main Still Picture Profile,", *IEEE TCSVT*, vol. 25, no. 5, pp. 790–797, 2015.

[3] J. Lee and T. Ebrahimi, "Perceptual video compression: a survey,", *IEEE JSTSP*, vol. 6, no. 6, pp. 684–697, 2012.

[4] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *PIEEE*, vol. 101, no. 9, pp. 2025–2043, 2013.

[5] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE TIT*, vol. 20, no. 4, pp. 525–536, 1974.

[6] I. Höntsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," *IEEE TIP*, vol. 11, no. 3, pp. 213–222, 2002.

[7] D. M. Chandler and S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *IEEE TIP*, vol. 14, no. 4, pp. 397–410, 2005.

[8] Z. Liu, L. J. Karam, and A. B. Watson, "Jpeg2000 encoding with perceptual distortion control," *IEEE TIP*, vol. 15, no. 7, pp. 1763–1778, 2006.

[9] K.-h. Park and H. Park, "Region-of-interest coding based on set partitioning in hierarchical trees," *IEEE TCSVT*, vol. 12, no. 2, pp. 106–113, 2002.

[10] L. Liu and G. Fan, "A new jpeg2000 region-of-interest image coding method: partial significant bitplanes shift," *IEEE SPL*, vol. 10, no. 2, pp. 35–38, 2003.

[11] P. G. Tahoces, J. R. Varela, M. J. Lado, and M. Souto, "Image compression: Maxshift roi encoding options in jpeg2000," *CVIU*, vol. 109, no. 2, pp. 139–145, 2008.

[12] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE TIP*, vol. 19, no. 1, pp. 185–198, 2010.

[13] S. S. Channappayya, A. C. Bovik, and R. W. Heath Jr, "Rate bounds on ssim index of quantized images," *IEEE TIP*, vol. 17, no. 9, pp. 1624–1639, 2008.

[14] D. Schonberg, G. J. Sullivan, S. Sun, and Z. Zhou, "Perceptual encoding optimization for jpeg xr image coding using spatially adaptive quantization step size control," in *SPIE Optical Engineering+ Applications.* International Society for Optics and Photonics, 2009, pp. 74 430M–74 430M.

[15] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE TIP*, vol. 15, no. 11, pp. 3440–3451, 2006.

[16] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE TIP*, vol. 20, no. 5, pp. 1185–1198, 2011.

[17] B. Li, H. Li, L. Li, and J. Zhang, "λ domain based rate control for high efficiency video coding," *IEEE TIP*, vol. 23, no. 9, Sep. 2014.

[18] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE SPM*, vol. 15, no. 6, pp. 74–90, 1998.

[19] X. W. Marta Karczewicz, "Intra frame rate control based on SATD," *Document: JCTVC-M0257, Joint Collaborative Team on Video Coding*, Apr. 2013.

[20] S. Fan, "A new extracting formula and a new distinguishing means on the one variable cubic equation," *Nat. Sci. J. Hainan Teach. Coll*, vol. 2, pp. 91–98, 1989.

[21] X. Mai, R. Yun, and W. Zulin, "Learning to predict saliency on face images," *ICCV*, 2015.

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