Supporting Document

Saliency-Guided Complexity Control for HEVC Decoding

Ren Yang, *Student Member, IEEE*, Mai Xu, *Member, IEEE*, Zulin Wang and Xiaoming Tao

I. Eye-Tracking Experiment

We conducted the eye-tracking experiment to obtain the gournd-truth fixations of all 15 sequences of the JCT-VC database (Classes A-D, except 10-bit sequences). In our eye-tracking experiment, 32 subjects (18 males and 14 females), with either corrected or uncorrected normal eyesight, have participated in viewing all 15 sequences. Note that only 2 among 32 subjects were experts, who worked on the research field of saliency detection. The other 30 subjects did not have any background in saliency detection, and they were native to the purpose of the eye tracking experiment. Then, a Tobii TX300 eye tracker, integrated with a 23-inch LCD displaying screen, was used to track the eye movement at a sample rate of 300Hz. All subjects were seated on an adjustable chair at a distance of around 60 cm from the screen of the eye tracker. Before the experiment, subjects were instructed to perform the 9-point calibration for the eye tracker. During the experiment, each sequence was presented in a random order, followed by a 2-second black image for a drift correction. All subjects were asked to free-view each sequence. Finally, 183,507 fixations of 32 subjects were collected for all 15 sequences after the experiment. The database of those fixations is available at https://github.com/SGCCmaterials/Fixations.git.

II. Proof for Lemmas 3-4 and Proposition 5

A. Proof for Lemma 3

First, upon (23), $I = \sum_{n=1}^{N} f_n$ holds. Defining $M = \sum_{n=1}^{N} f_n^{'}$, we can turn (24) to

$$a \cdot \left(\sum_{n=1}^{N} w_n \cdot f_n \right) + I \cdot b = a \cdot \left(\sum_{n=1}^{N} w_n \cdot f_n^{'} \right) + M \cdot b. \tag{32}$$

If $I \geq M$, then $I \cdot b \geq M \cdot b$ exists due to $b > 0$. Given $I \cdot b \geq M \cdot b$ and (32), we can obtain (25) because $a > 0$. Therefore, for the proof of (25), we only need to prove $I \geq M$.

Next, we prove $I \geq M$ by contradiction as follows. In the case of $I < M$, we have $\sum_{n=1}^{N} f_n < \sum_{n=1}^{N} f_n^{'}$. Because $f_n = 1$ holds if and only if w_n ($\in [0, 1]$) belongs to the smallest I values in $\{w_n\}_{n=1}^{N}$, the following inequality exists,

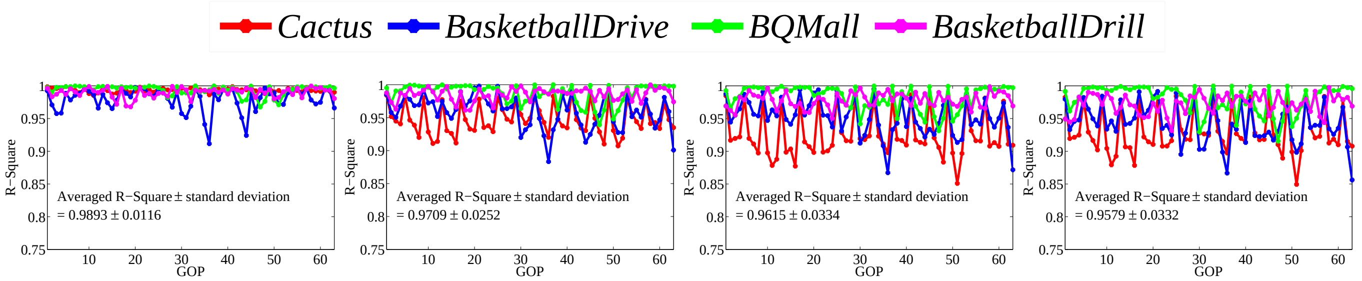
$$\sum_{n=1}^{N} w_n \cdot f_n < \sum_{n=1}^{N} w_n \cdot f_n^{'}. \tag{33}$$

When $I < M$ and $b > 0$, it is obvious that $I \cdot b < M \cdot b$ holds. Then, given $a > 0$, we can obtain

$$a \cdot \left(\sum_{n=1}^{N} w_n \cdot f_n \right) + I \cdot b < a \cdot \left(\sum_{n=1}^{N} w_n \cdot f_n^{'} \right) + M \cdot b. \tag{34}$$

However, (34) contradicts with (32). Hence, the assumption of $I < M$ dose not hold, such that $I \geq M$ can be proved. Finally, the inequality of (25) is achieved.

This completes the proof of Lemma 3. ■



(a) QP = 22 (b) QP = 27 (c) QP = 32 (d) QP = 37

Fig. 1. The R-square values for the second-order polynomial regression of $\sum_{n=1}^{N_t} \widetilde{w}_n = k \cdot N_t^2$. Note that the R-square values of the 3rd frames in each GOP are shown.

B. Proof for Lemma 4

First, $\sum_{n=1}^{N} g_n = N_1 + 2N_2 + 3N_3$ holds, since N_0, N_1, N_2 and N_3 are the numbers of CTUs at $g_n = 0, 1, 2, 3$. As such, the constraint terms of (22-b) and (28) are equivalent. Next, constraint term $\frac{1}{N} \cdot c \cdot (N_1 + 2N_2 + 3N_3)$ of (22-b) is independent of w_n. Hence, larger g_n should correspond to smaller w_n to make $\sum_{n=1}^{N} w_n \cdot (h_1 \cdot g_n^3 + h_2 \cdot g_n^2 + h_3 \cdot g_n)$ minimal. This way, for each combination of N_0, N_1, N_2 and N_3, the optimal solution satisfies $\forall w_n \leq w_{n'}, g_n \geq g_{n'}$. Then, the values of N_0, N_1, N_2 and N_3 are the variables to be solved for the minimization problem of (22-b). Defining $\{\widetilde{w}_n\}_{n=1}^{N}$ as the ascending sort of $\{w_n\}_{n=1}^{N}$, the optimal solution $\{g_n\}_{n=1}^{N}$ towards (22-b) can be written as

$$g_n = \begin{cases} 3, & w_n \leq \widetilde{w}_{N_3} \\ 2, & \widetilde{w}_{N_3+1} \leq w_n \leq \widetilde{w}_{N_3+N_2} \\ 1, & \widetilde{w}_{N_3+N_2+1} \leq w_n \leq \widetilde{w}_{N_3+N_2+N_1} \\ 0, & \widetilde{w}_{N_3+N_2+N_1+1} \leq w_n \leq \widetilde{w}_{N}. \end{cases} \tag{35}$$

Upon (22-b) and (35), we can obtain (28). Note that $h_1 \cdot g_n^3 + h_2 \cdot g_n^2 + h_3 \cdot g_n = 1$ when $g_n = 3$, according to (16).

Finally, Lemma 4 can be proved. ■

C. Proof for Proposition 5

First, we apply our method of Section II to estimate the saliency values $\{w_n\}_{n=1}^{N}$ of all CTUs in the four training sequences (the same as Section IV) at four QPs (i.e., 22, 27, 32 and 37). Then, at each QP, the values of $\sum_{n=1}^{N_t} \widetilde{w}_n$ for all possible $N_t \in \{n\}_{n=1}^{N}$ are calculated and recorded at each frame of the four sequences. Recall that $\{\widetilde{w}_n\}_{n=1}^{N}$ is the ascending sorted set of saliency values $\{w_n\}_{n=1}^{N}$ for each frame. Next, we apply the second-order polynomial regression to each frame for modelling the relationship between N_t and $\sum_{n=1}^{N_t} \widetilde{w}_n$ in form of $\sum_{n=1}^{N_t} \widetilde{w}_n = k \cdot N_t^2$, where k is the second-order parameter of the regression for a video frame. Note that k is a constant within a video frame, despite being different across frames. The R-square values of such regression can be obtained across different frames of four training sequences. Fig. 1 shows R-square of $\sum_{n=1}^{N_t} \widetilde{w}_n = k \cdot N_t^2$ regression along with various frames. It is evident that the R-square values are above 0.85 for all frames at four QPs. In addition, the averaged R-square values and their standard deviations are 0.9893 ± 0.0116, 0.9709 ± 0.0252, 0.9615 ± 0.0334 and 0.9579 ± 0.0332 at QP = 22, 27, 32 and 37. Thus, $\sum_{n=1}^{N_t} \widetilde{w}_n$ can be well approximated by $k \cdot N_t^2$.

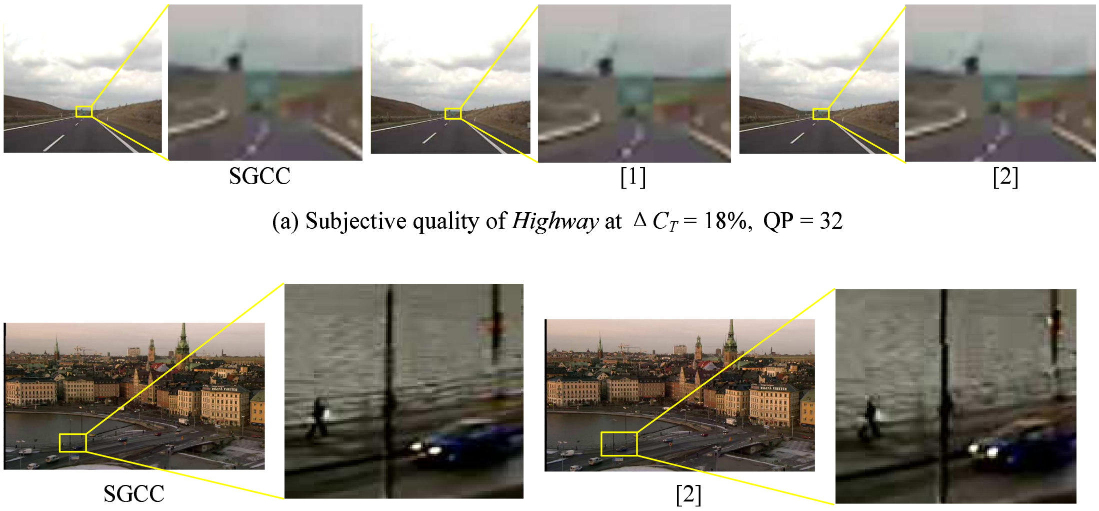
Finally, Proposition 5 can be proved. ■

III. DMOS Experiment

In our experiment, the DMOS test was conducted to rate subjective quality of the decoded sequences, by the means of Single Stimulus Continuous Quality Score (SSCQS), which is processed by Rec. ITU-R BT.500. The total number of subjects involved in the test is 12, consisting of 7 males and 5 females. Note that, these subjects are totally different from those selected for the eye-tracking experiment. Here, a Sony BRAVIA XDVW600 television, with a 55-inch LCD displaying screen, was utilized to display the decoded sequences. The viewing distance was set to be approximately four times of the video height for rational evaluation. During the test, sequences were displayed in random order. After viewing each decoded sequence, the subjects were asked to rate the sequence. The rating score includes excellent (100-81), good (80-61), fair (60-41), poor (40-21), and bad (20-1). Finally, DMOS was calculated upon the scores of decoded sequences rated by 12 subjects. As a result, DMOS value of each decoded sequence can be obtained to measure the difference of subjective quality between sequences decoded by original HEVC and by HEVC with our SGCC approach or other conventional approaches [23] [24].

IV. Subjective quality for sequences without dominated objects

The subjective quality of the sequences without dominated objects, which are decoded by our SGCC, [23] and [24] approaches, is shown in Figure 2. As shown, in *Highway*, [23] and [24] have obvious blocky artifacts on the road ahead, i.e., Region of Interests (ROI). In *Stockholm*, [24] incurs severe distortion on the pedestrian and the pole. In the contrast, our SGCC approach produces less artifacts than [23] and [24], especially at ROI. In a word, the subjective quality of our SGCC approach outperform [23] and [24] on sequences without dominated object.



(a) Subjective quality of *Highway* at $\Delta C_T = 18\%$, QP = 32

(b) Subjective quality of *Stockholm* at $\Delta C_T = 23\%$, QP = 32

Fig. 2. Subjective quality of videos without dominated object.