SALIENCY BASED PERCEPTUAL HEVC

Yiming Li, Weihang Liao, Junming Huang, Da He, and Zhenzhong Chen

Wuhan University, Wuhan, Hubei, 430079, P. R. China

ABSTRACT

Saliency represents the probability of human attention over the image therefore is important in understanding the importance of different areas in the image. In this paper, we consider how to utilize the saliency information in HEVC video coding. The graph-based visual saliency is incorporated with the quantization control in HEVC to reduce the bit rate of compressed video by using larger quantization parameter for the coding unit which has lower probability of attention. As shown in our experiments, the proposed method achieves up to 12% bit rate reduction without perceptual quality loss.

Index Terms— saliency, HEVC, quantization parameter

1. INTRODUCTION

With the development of Internet and wireless technologies, more and more video related applications have been developed, such as Video on Demand (VoD), video conference, video sharing, etc. However, with the rapid grown of video over network, the limited network bandwidth becomes the bottomneck. Therefore, high efficient video compression technology is highly desired.

High Efficiency Video Coding (HEVC) [1], as a successor to H.264/MPEG-4 AVC (Advanced Video Coding), was jointly developed by the ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG) as ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T H.265 [2]. HEVC doubles the data compression performance compared to H.264/MPEG-4 AVC by reducing the bit rate by half whilest maintain the same level of video quality [3] [4].

Compared with H.264/AVC, HEVC standard has some new features, such as quadtree structure of the coding unit, sample adaptive offset (SAO), advanced motion vector prediction (AMVP), etc [1] [5]. In HEVC, a frame is divided into Coding Tree Units (CTUs) which can use a large block structure of up to 64×64 pixels. The block can be divided into coding units (CUs) continually by using quadtree syntax of the CTU [6] [7]. Thus, HEVC may adapt to high resolution video coding. A CU can be further split into PUs and TUs, where HEVC defines PU (Prediction Unit) for prediction coding and TU (Transform Unit) for transform. In interpicture prediction, HEVC allows advanced motion vector prediction (AMVP) to improve coding efficiency while a merge mode for motion vector coding is used, as well [8]. In addition, HEVC contains several techniques to make it more parallel-friendly than H.264/AVC [9][10]. In addition, sample adaptive offset (SAO) is added in HEVC to reconstruct the signal [1].

With the study of human visual system (HVS) and the development of the prediction capacity in human visual perception, perceptual video coding attracts more and more attention. In fact, it is well known that the distortion introduced by quantization in lossy coding is not uniformly perceived, but it rather varies due to some HVS masking effects [11] [12]. Perception model can help us to improve the video coding efficiency [13] [14]. In [15], perceptual video coding is classified into three categories, vision-model based approach, signal-driven approach, and hybrid approach. In this paper, a saliency information based perceptual HEVC is proposed to determine the quantization parameter of CU, as shown in Fig. 1. The word saliency describe some part of a scene, which can be a region or an object, is somehow more stand out than their neighborhoods [16]. The first definition of saliency map was given by Koch and Ullman [17]. It is a measurement of the visual attraction of each point in a scene. By applying a saliency detection model we can get a saliency map, describing how the features in the image attract peoples attention, [18] which can be used to measure the visual importance region in a frame of a video. Thus with a saliency information, the proposed method allows the encoder to select proper quantization parameter for a CU based on its visual importance and improve the coding efficiency by achieving bit rate reduction without visual quality loss.

The rest of the paper is organized as follows. Saliency based perceptual HEVC is described in Section II, which includes the framework of HEVC, the method to get saliency input and the quantization control in HEVC. Section III provides subjective test environment and implementation details while Section IV concludes the paper.

2. SALIENCY BASED PERCEPTUAL HEVC

2.1. Framework

The framework of the saliency based HEVC is shown in Fig.1. The saliency information is inputed for HEVC to determine the quantization parameter of CU. For instance, a

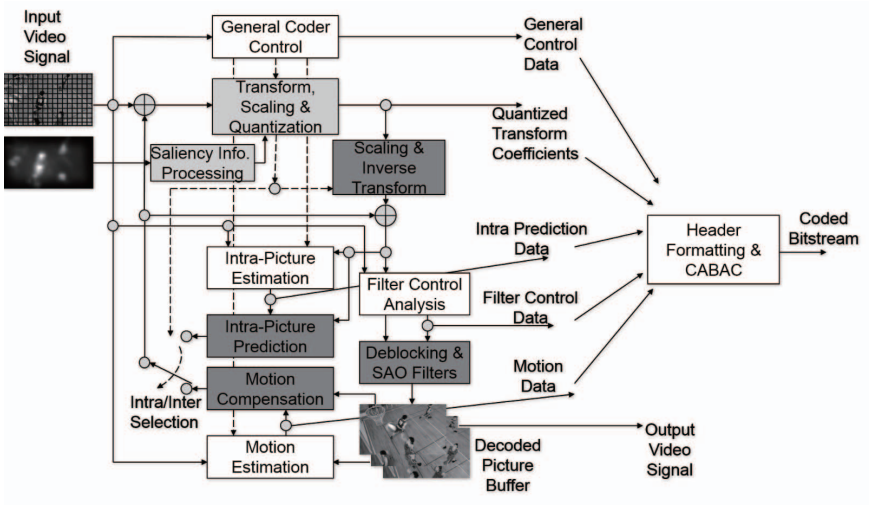

Fig. 1. saliency based perceptual HEVC framework.

smaller QP will be used if the CU is important as indicated by its corresponding saliency information.

2.2. Saliency input

Saliency refers to the indication of the probability of human attention according to the visual stimulus. Here we use the GBVS (Graph-Based Visual Saliency) [2] to get the initial saliency map. A Gaussian function is used as to refine the initial saliency map. Therefore, we obtain the saliency map which can distinguish the CUs with different probabilities of attracting human attention.

Our saliency calculation algorithm consists of three steps:

- Form an activation map

  For each given frame I, we extract feature vectors to generate a feature map M, this step is done by using biologically inspired filters. Then with each feature map M, we need to calculate an activation map A. Intuitively, if a pixel $I(i,j)$ in the original frame I is unusual in neighborhood, the corresponding $A(i,j)$ will have a high value in activation map A. In consideration of the influence of the surrounding pixels, the value should be calculated in the region around (i,j). Thus we compute the level of unusual as

$$A(i,j) = -\log(p(i,j)) \tag{1}$$

$$p(i,j) = PrM(i,j)|neighborhood \tag{2}$$

- Normalizing and Combination

  After we get activation map for each frame, we use a Markovian algorithm to concentrating it into a few key locations. For each activation map A which need normalization, we construct a graph GN with n2 nodes labelled with indices from A. For each node (i,j) and node (m,n), if they are connected we introduce an edge from (i,j) to (m,n) with weight:

$$W((i,j),(m,n)) = A(m,n) \times F(i-m, j-n) \tag{3}$$

$$F(a,b) \triangleq \exp(-\frac{a^2+b^2}{2\eta^2}) \tag{4}$$

  where η is a free parameter of the algorithm. By using a Markov chain, we can compute the equilibrium distribution over the nodes. For each frame, we compute a set of activation maps, then we dispose each activation map with normalization algorithm, then simply sum over the feature channels to get the saliency map.

- Postprocessing Typically, by considering that people focus on the center when viewing an image or video, we use a Gaussian function to refine our results as:

$$S(i,j) = S(i,j) \cdot \exp(\frac{(i-CenX)^2 + (j-CenY)^2}{-2 \cdot \sigma^2}) \tag{5}$$

$$\sigma = \sqrt{\frac{CenX^2 + CenY^2}{-2 \cdot \log 0.5}} \tag{6}$$

  Where CenX, CenY is the center pixel of the frame.

2.3. Quantization Control

In HEVC, each frame can be divided into CUs (coding unit), while quantization parameter (QP) and quantization step is related to each CUs bit rate and definition. Quantization is lossy, when we enlarge quantization step, the bit rate can be reduced relatively. The range of QP is from 0 to 51, and there are exponent relation betweenin QP and quantization step, in

addition, the lager value of QP, the less amount of bit rate and larger distortion of the image. So, we can reduce the bit rate of the CU far away from the focus center by coding these CUs with lager QP value. In this way, we may achieves much bit rate reduction without perceptual quality loss. With its saliency pixel value $y(i,j)$ and the average value x of the whole frame, the CU's QP can be determined according to its importance in terms of visual properties. Based on the saliency information, the CU with lower visual attention probability will be given larger QP value. Let α and β be two predefined parameters, which are the thresholds that control QP adaptively, e.g., $\alpha = 0.5 \times 2x$ and $\beta = 0.8 \times 2x$, we have

$$QP(i,j) = \begin{cases} QP_0 + m, & y(i,j) \leq \alpha \\ Int[QP_0 + m + \frac{m}{\alpha - \beta} \\ \times (y(i,j) - \alpha)], & \alpha < y(i,j) \leqslant \beta \\ QP_0, & y(i,j) > \beta \end{cases} \quad (7)$$

where QP_0 is the initial quantization coefficient, n is the maximum modification of QP, and m is the minimal modification of QP.

Algorithm 1 Quantization Control Algorithm

input:

image of each frame of the sequence: I, saliency map of the sequence: S, parameter to control threshold:TH_0,TH_1, and QP_0, n, m, which has introduced above.

for all $CU(i,j) \in I$ **do**

y(i,j)←get mean value of (i,j) block in S

Average←get mean value of the whole part in S

compute: α←2· TH_0·Average

β←2· TH_1·Average

$$QP(i,j) = \begin{cases} QP_0 + m, & y(i,j) \leq \alpha \\ Int[QP_0 + m + \frac{m}{\alpha - \beta} \\ \times (y(i,j) - \alpha)], & \alpha < y(i,j) \leqslant \beta \\ QP_0, & y(i,j) > \beta \end{cases} \quad (8)$$

QP(i,j) =min(QP(i,j),51)

QP(i,j) =max(QP(i,j),0)

end for

output:

QP(i,j)

3. EXPERIMENTAL RESULT

Subjective tests have been conducted to evaluate the performance of the proposed approach for HEVC coding applications in a typical laboratory viewing environment with normal lighting. The original HM method in HEVC is used for comparisons. The display system was a 30 inch DELL U3014 monitor. The protocol we used is the double-stimulus continuous quality scale (DSCQS) protocol recommended in Rec. ITU-R BT.500 [3]. The assessors were carefully introduced the assessment methods, the quality or impairment factors such as occur, grading scales, and timing. We used test sequences from class A to class F. Class A's resolution is 2560×1600 and frame rate is 30 fps. Class B's resolution is 2560×1600 and frame rate is 24fps. Class C's resolution is 832×480 and frame rate 50 or 60fps. Class D's resolution is 416×240 and frame rate 30 or 50fps. Class E's resolution is 1280×720 and frame rate 60 fps. Class F's resolution is 352×288 and frame rate 30 fps. Each sequence is coded with four quantization parameter values: 22, 27, 32, and 37, respectively. The proposed method is compared to the HM (revision 3517).

Fig. 2. The original frame from test video BQMall.

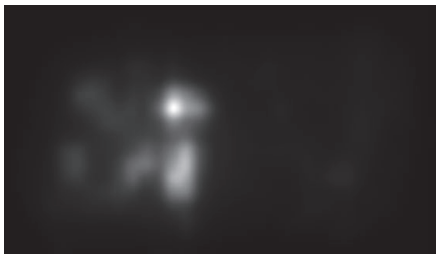

Fig. 3. The saliency map result.

The DSCQS protocol has been used to evaluate the quality of a pair of videos. Sixteen people with normal (or corrected to normal) vision have been participated in this test, nine female and seven male. The procedure is to make a comparison of stimulus video A and B, each of which is 10 second. Stimulus video A is either the HEVC HM result, or the result from the proposed method. Stimulus video B is the other result accordingly. Therefore, the two stimulus videos from two coders are randomly presented in the order. We showed video A and B in turn. The duration between two sequences is 3 s. Each pair of videos is displayed two times. The second time is the voting time when the observers were asked to vote. The comparison scales for the DSCQS protocol range from -3 to 3, where 0 means video B is with the same quality as A, -3 means video B is much worse than A, -2 means video B is worse than A and -1 means slightly worse, 3 means

video B is much better than A, 2 means video B is better than A and 1 means slightly better. Table II provides the comparisons of the subjective test results. The confidence interval is 95%. As recommended in Rec. ITU-R BT.500 [3], the Kurtosis coefficient, $\beta 2$, was calculated to determine whether an observer should be rejected. The results show that the bit rate is reduced upto 12% and the mean comparison scales are negligible.

4. SUMMARY

In this paper, a saliency information based perceptual HEVC is proposed to determine the quantization parameter of CU. With the study of human visual system, we change the quantization parameter of different coding union basis of human's attention adaptively. As shown in subjective tests, the proposed method achieves up to 12% bit rate reduction without perceptual quality loss.

5. REFERENCES

[1] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard,"IEEE Trans. Circuits and Systems for Video Technology, 22.12 (2012): 1649-1668.

[2] B. Bross, W. Han, G. Sullivan, J. Ohm, and T. Wiegand, "High Efficiency Video Coding (HEVC) Text Specification Draft 9,"JCTVC-K1003, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), Oct. 2012.

[3] F. Bossen, B. Bross, K. Suhring, and D. Flynn, "HEVC Complexity and Implementation Analysis,"IEEE Trans. Circuits Syst. Video Technol., 22.12 (2012), 1685-1696.

[4] T. Wiegand, J. R. Ohm, G. J. Sullivan, W. J. Han, R. Joshi, T. K. Tan and K. Ugur, "Special section on the joint call for proposals on High Efficiency Video Coding (HEVC) standardization,"IEEE Trans. Circuits and Systems for Video Technology, 20.12 (2010): 1661-1666.

[5] O. Le Meur, "Video compression Beyond H. 264, HEVC,"(2011).

[6] Y. Yuan, I. K. Kim, and X. Zheng, "Quadtree based nonsquare block structure for inter frame coding in high efficiency video coding,"IEEE Trans. Circuits and Systems for Video Technology, 22.12 (2012): 1707-1719.

[7] D. Marpe, H. Schwarz, and S. Bosse, "Video compression using nested quadtree structures, leaf merging, and improved techniques for motion representation and entropy coding,"IEEE Trans. Circuits and Systems for Video Technology, 20.12 (2010): 1676-1687.

[8] J. L. Lin, Y. W. Chen, and Y. P. Tsai, "Motion vector coding techniques for HEVC,"IEEE 13th International Workshop on Multimedia Signal Processing (MMSP), 2011.

[9] C. Yan, Y. Zhang, and F. Dai, "Highly parallel framework for hevc motion estimation on many-core platform,"IEEE Data Compression Conference (DCC), 2013.

[10] M. Zhou, "AHG10: Configurable and CU-group level pa rallel merge/skip,"JCTVC-H0082, Feb. 2012.

[11] M. Naccari and F. Pereira, "Advanced H. 264/AVC-based perceptual video coding: Architecture, tools, and assessment,"IEEE Trans. Circuits and Systems for Video Technology, 21.6 (2011): 766-782.

[12] H. R. Wu and K. R. Rao, "Digital Video Image Quality and Perceptual Coding,"Boca Raton, FL: CRC Press, Nov. 2005, p. 640.

[13] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency,"proceedings of Neural Information Processing Systems (NIPS), 2006.

[14] Z. Chen and C. Guillemot, "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model,"IEEE Trans. Circuits and Systems for Video Technology, 20.6 (2010): 806-819.

[15] Z. Chen, W. Lin, and K. N. Ngan, "Perceptual video coding: challenges and approaches,"IEEE International Conference on Multimedia and Expo (ICME), 2010.

[16] A. Borji and L. Itti, "State-of-the-art in visual attention modeling,"IEEE Trans. Pattern Analysis and Machine Intelligence, 35.1 (2013): 185-207.

[17] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry,"Matters of Intelligence. Springer Netherlands, 1987. 115-141.

[18] A. Treisman and G. Gelade, "A feature-integration theory of attention,"Cognit. Psychol., vol. 12, no. 1, pp. 97136, 1980.

[19] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency,"proceedings of Neural Information Processing Systems (NIPS), 2006.

[20] ITU-T. BT500-11, "Methodology for the subjective assessment of the quality of television pictures."(2002).

Table 1. Result of DSCQS Tests for Video Coding.

Class	Sequence Name	QP=22						QP=27					
		Bit rate (kbps)		PSNR(Y) (dB)		Mean Comparison	Bit rate	Bit rate (kbps)		PSNR(Y) (dB)		Mean Comparison	Bit rate
		HM	Proposed	HM	Proposed	Scale	Saving	HM	Proposed	HM	Proposed	SCale	Saving
A	PeopleOnStreet	32819.76	30128.88	40.18	39.75	-0.313	8.20%	15718.12	14861.98	37.16	36.87	0.133	5.45%
B	Kimono	4735.4	4127.12	41.6	41.28	0	12.85%	2164.1	1946.73	39.73	39.37	-0.133	10.04%
B	ParkScene	7413.76	6402.65	40.05	39.47	0.188	13.64%	3179.41	2841.71	37.51	36.95	-0.2	10.62%
C	BQMall	3833.12	3433.26	40.23	39.76	0.063	10.43%	1823.78	1681.33	37.74	37.26	0.267	7.81%
C	PartyScene	7105.71	6278.13	38.31	37.48	0.125	11.65%	3298.06	2993.13	34.8	34.12	0.067	9.25%
D	RaceHorses	1198.35	1075.67	39.47	38.71	-0.438	10.24%	587.79	540.04	35.74	35.16	-0.133	8.12%
D	BasketballPass	1528.38	1383.63	40.71	39.96	-0.125	9.47%	767.81	707.09	36.95	36.43	-0.267	7.91%
E	MoblieCalendar	9449.6	7758.82	38.65	38.19	0.063	17.89%	3001.89	2634.25	36.81	36.25	0.133	12.25%
E	City	10722.5	9144.42	39.3	38.69	0.25	14.72%	3366.12	2975.71	36.75	36.23	0.2	11.60%
F	Akiyo	141.23	128.12	44.05	43.37	0.125	9.28%	74.61	69.05	41.5	40.77	-0.267	7.45%
F	HallMonitor	472.78	402.95	40.3	39.84	-0.125	14.77%	163.58	148.97	38.16	37.7	-0.133	8.93%
F	Silent	433.33	382.62	40.84	39.72	0.063	11.70%	217.78	196.14	37.28	36.37	0.133	9.93%
	Average						**12.07%**						**9.11%**

Class	Sequence Name	QP=32						QP=37					
		Bit rate (kbps)		PSNR(Y) (dB)		Mean Comparison	Bit rate	Bit rate (kbps)		PSNR(Y) (dB)		Mean Comparison	rate Saving
		HM	Proposed	HM	Proposed	Scale	Saving	HM	Proposed	HM	Proposed	Scale	Saving
A	PeopleOnStreet	8250.61	7903.99	34.21	33.97	0.133	4.20%	4620.85	4480.06	31.44	31.26	0.4	3.05%
B	Kimono	1054.08	956.46	37.42	37.02	0.133	9.26%	533.55	486.79	35.03	34.61	0.133	8.76%
B	ParkScene	1450.55	1322.24	34.91	34.37	-0.267	8.85%	670.73	615.31	32.39	31.89	-0.533	8.26%
C	BQMall	931.95	872.45	35.02	34.54	0.133	6.38%	499.25	469.8	32.28	31.81	0.2	5.90%
C	PartyScene	1588.67	1458.07	31.66	31.01	0.133	8.22%	770.04	707.3	28.75	28.09	-0.267	8.15%
D	RaceHorses	285.16	264.47	32.34	31.89	-0.2	7.26%	140.97	131.37	29.54	29.22	-0.067	6.81%
D	BasketballPass	379.51	355.63	33.59	33.21	-0.133	6.29%	197.59	186.44	30.8	30.47	-0.133	5.64%
E	MoblieCalendar	1402.31	1292.84	34.58	33.88	-0.2	7.81%	728.9	674.56	31.87	31.06	-0.133	7.45%
E	City	1480.17	1355.64	34.45	33.85	-0.067	8.41%	702.33	654.41	31.71	31.1	-0.133	6.82%
F	Akiyo	41.85	39.64	38.66	37.93	0.067	5.28%	25.12	23.78	35.67	34.84	-0.267	5.34%
F	HallMonitor	78.19	74.07	35.8	35.23	-0.133	5.27%	43.86	41.42	33.15	32.45	0.133	5.57%
F	Silent	109.92	100.34	34.04	33.34	-0.267	8.71%	55.33	51.06	31.24	30.71	-0.267	7.72%
	Average						**7.16%**						**6.62%**