RECENT ADVANCES IN PERCEPTUAL H.265/HEVC VIDEO CODING

(Invited Paper)

Zhenzhong Chen and Yiming Li

School of Remote Sensing and Information Engineering, Wuhan University, China

ABSTRACT

H.265/High Efficiency Video Coding (HEVC), the latest video coding standard developed jointly by ITU-T and by ISO/IEC in the joint collaborative team on video coding (JCT-VC), has shown significant objective compression improvements over the last video coding standard H.264/AVC. After the establishment of H.265/HEVC, more and more attention has been paid on how to further improve the visual quality such that better coding efficiency could be achieved. In this paper, we review the progress of subjective optimization of H.265/HEVC based video coding systems and summarize the recent advances of perceptual H.265/HEVC video coding. Referring to our earlier work, we classify the current perceptual H.265/HEVC video coding technologies into two categories, vision-model based approach and signal-driven approach. Moreover, due to the new technologies such as Coding Tree Unit (CTU) structure and Sample Adaptive Offset (SAO). The future research directions for perceptual H.265/HEVC video coding are also discussed.

***Index Terms*—** H.265/HEVC, perceptual coding, subjective optimization

1. INTRODUCTION

H.265/High Efficiency Video Coding (HEVC), developed jointly by ITU-T SG 16 Q.6, also known as the Video Coding Experts Group (VCEG), and by ISO/IEC JTC 1/SC 29/WG 11, also known as the Moving Picture Experts Group (MPEG) in the joint collaborative team on video coding (JCT-VC) in 2013 [1], succeeds H.264/AVC as the next generation video compression standard. Compared to H.264/AVC or other existing standards, H.265/HEVC shows significant performance improvements. Although it follows the hybrid coding structure, it has adopted some novel coding tools such as quadtree coding structure, sample adaptive offset (SAO), advanced motion vector prediction, etc. As a new component, coding tree unit (CTU) replaces 16x16 pixel macroblocks in earlier standards. In H.265/HEVC, the picture is divided into CTUs that the size of a CTU varies from 64x64 to 16x16. The division of CTU adapts to the content and shows important contributions to high resolution videos. Prediction units (PUs) and transform units (TUs) are also included in a CU, for prediction coding and for transform, respectively. In addition, sample adaptive offset (SAO) is designed in H.265/HEVC which can improve the quality of reconstructed picture.

Though H.265/HEVC is the most advanced video compression standard and has many advanced coding tools. It has some considerations on the subjective optimization based on the characteristics of human visual system (HVS), i.e., deblocking. Better utilizing the advantages of the human visual system can further remove visual redundancy therefore improve the compression efficiency. Based on H.265/HEVC video compression framework, some perceptual optimization approaches have been developed, e.g., combination of perceptual features with the new coding tools such as CTU [2] or SAO [3].

The rest of the paper is organized to describe these recent advances in perceptual H.265/HEVC video coding technology in details. In Section 2, a brief introduction of H.265/HEVC is provided. Perceptual H.265/HEVC video coding is summarized in Section 3. An example method is described in Section 4 to demonstrate the advantages of perceptual optimization for H.265/HEVC. In Section 5, we conclude the paper.

2. INTRODUCTION OF H.265/HEVC

H.265/HEVC was established in January 2013. Compared to the earlier standard H.264/AVC, H.265/HEVC has some new features, such as quadtree structure of the coding unit, sample adaptive offset (SAO), advanced motion vector prediction (AMVP), etc [1]. In H.265/HEVC, a frame is divided into Coding Tree Units (CTUs) which can use a large block structure of up to 64x64 pixels. The block can be divided into coding units (CUs) continually by using quadtree syntax of the CTU. Thus, H.265/HEVC may adapt to high resolution video coding. A CU can be further split into PUs and TUs, where H.265/HEVC defines PU (Prediction Unit) for prediction coding and TU (Transform Unit) for transform. In inter-picture prediction, H.265/HEVC allows advanced mo-

This work was supported in part by National Natural Science Foundation of China (No. 61471273).

tion vector prediction (AMVP) to improve coding efficiency while a merge mode for motion vector coding is used. In addition, sample adaptive offset (SAO) is added in H.265/HEVC to reconstruct the signal. H.265/HEVC also contains several techniques to make it more parallel-friendly [1]. Based on these new techniques, H.265/HEVC doubles the compression ratio compared to H.264/ AVC at the same level of visual quality.

3. PERCEPTUAL H.265/HEVC VIDEO CODING

With the progress of research on human visual system and development on video compression systems, new perceptual optimization modules have integrated into the video coding systems. With the establishment of the latest video coding standard, H.265/HEVC, some attempts have been made to optimize the subjective quality of the H.265/HEVC video coding systems. To summarize different perceptual video coding algorithms applied on the H.265/HEVC framework, we refer to [4] to classify them into two categories, vision-model based approach and signal-driven approach.

3.1. Vision-model based approaches

3.1.1. ROI based video coding

It is well known that when people watch a video or picture, they may only pay attention to particular region or object in the visual scene instead of the whole frame or image. Therefore, Region of interests (ROI) or object of interests (OoI) based perceptual video coding has been studied. However, as H.265/HEVC introduces the new quadtree coding structure, it brings new challenges in the design and implementation of the ROI based H.265/HEVC. A bit allocation scheme based on hierarchical perception model of face has been proposed in [2, 5]. Considering the eyes, the mouth, and the other face areas are of different levels of interests, different weights for different face regions are set. In this ROI video coding scheme, the larger weight of the region, the greater depth LCU for split is allowed and the finer quantization parameter (QP) is used, that is called weight-based unified rate-quantization (URQ) scheme instead of pixel-based URQ scheme [2].
ROI may not only refer to human face in the picture, but the general foreground region in contrast of background region. Liang et al. [6] propose a scene-aware perceptual video coding by scene reconstruction to recognition the foreground area and the background area. The structure from motion (SFM) technology is used to reconstruct the 3D point of each scene of frames and then K-means algorithm is employed for clustering foreground and background. Moreover, to protect objects boundaries, the authors propose to use the corresponding distance information to adjust the QP value to obtain the better subjective quality.

3.1.2. Attention based video coding

The techniques for computational visual attention model have been extensively explored in recent years. When human observe the scenes, the visual attention is affected by two manners, bottom-up visual attention and top-down visual attention. Top-down attention is task-dependent, volition-controlled and knowledge based, while bottom-up attention is task-independent and stimulus based. Itti et al. [7] proposed a computational visual attention model based on the feature integration theory to analyze the attention region in the visual scene. After then, people propose different attention models and utilize them in perceptual image or video coding.
Taking these research outputs into the H.265/HEVC framework, Li et al. [8] utilize the saliency map calculated from the computational visual attention model to adjust bits allocation for better perceptual quality. Milani et al. [9] use object detection algorithm to generate saliency metric and optimize the object edge bits allocation. As shown in these two methods, an attention based perceptual video coding typically consists of following steps: the first step is pre-generating a saliency map/metric of each frame by a saliency model, and then, add these saliency map/metric into the video compression loop to adaptively allocate bit rate to different coding unit according to the saliency map/matric. More specifically, the idea is to increase or decrease the quantization steps based on the probabilities of the coding units which are indicated by the saliency map/metric. In this way, we can preserve the fine details of region of attention (ROA) and reduce the bit rate by saving bits from other regions.

3.2. Signal-driven approaches

3.2.1. Perceptual metric based video coding

The structural similarity index (SSIM) is a better evaluation for perceptual video quality compared with peak signal to noise ratio (PSNR) or the sum of square error (SSE). So in some work, SSIM is used in rate-distortion optimization (RDO). Yeo et al. [10] propose an SSIM-based RDO for H.265/HEVC. They calculate the conversion relation between SSIM and MSE and then use SSIM value to replace MSE in RDO. In this way, they claim that they can obtain better subjective video performance as SSIM can better measure the visual quality of the reconstructed video. There is also a work that performs SSIM-inspired RDO in H.265/HEVC based on divisive normalization [11].

3.2.2. Sensitivity based video coding

Just-Noticeable-Difference (JND) defines the smallest detectable difference between two signals therefore can be utilized to quantify the perceivable distortion in the noise contaminated image. Since the purpose of video coding is

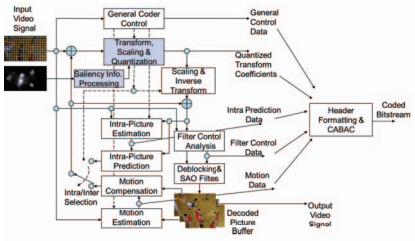

Fig. 1. Attention-based H.265/HEVC video coding [9].

to achieve highest perceptual quality, JND thresholds could hence be used to determine optimum quantization step sizes for different parts of video frame. There are some schemes utilize JND models [3, 12] for H.265/HEVC. Kim et al. [13] used the JND model for the transform skip mode (TSM) and the Transform non-skip models (non-TSMs) by adjust RDO parameters. For TSM, they reset the distortion of different size of transform unit (TU) block by the JND model of luminance masking (LM) effect, which size ranges from 4x4 to 32x32. For non-TSM, they use the JND model of temporal masking (TM) effect, contrast masking (CM) effect and generate contrast sensitivity function (CSF) besides LM effect. Yang et al. [3] propose an SAO RDO method based on JND model which can largely reduce the computational complexity by introducing JND model in SAO. These two proposed perceptual video coding approaches use the JND model in two different modules in H.265/HEVC frames, i.e., one is for transform&quantization and the other one is for SAO. Based these approaches, the perceptual quality of H.265/HEVC can be improved.

3.2.3. Texture based video coding

Ndjiki-Nya et al. [12] has utilized content based video coding in video coding based on texture analysis and synthesis. The work has been further extended for H.265/HEVC [14]. Based on this video coding technology, the encoder can skip some regions since the texture has been analyzed and stored while decoder can refer to texture synthesis to reconstruct the region.

3.2.4. Temporal optimization based video coding

In addition to above signal-driven approaches, Adzic et al. [15] propose a temporal perceptual coding. Based on the relationship between maximal spatial acuity and retinal velocity, a temporal visual acuity model is developed to improve the perceptual coding performance of H.265/HEVC by eliminating the need to signal coefficients based the frequency content and velocity.

(a) (b)

Fig. 2. The original video frame (ParkScene1920x1080 and the corresponding saliency map..

4. A PERCEPTUAL H.265/HEVC EXAMPLE

In this section, an example is provided to demonstrate the advantage of perceptual H.265/HEVC video coding. The example is oriented from the authors earlier attention-based H.265/HEVC [4]. The perceptual H.265/HEVC video coding framework is shown in Figure 1.

The saliency information is input for H.265/HEVC to determine the quantization parameter of CU. The higher the attention probabilities, the smaller the quantization parameters will be used. The original HM method is used for comparisons. A video sequence is coded by either the original HM 11.0 with QP value 32 or the perceptual approach where the QP value ranges from 32 to 37 according to the attention weights. Fig.2 shows the saliency map generated by the scene. The results of different attention regions or non-attention regions are shown in Fig 3 to illustrate the subjective quality comparisons.

For the attention region (a) and (c), as the same QP is used, there is no perceptual difference. For the non-attention regions (b) and (d), although the larger QP is used in the attention-based H.265/HEVC which results in higher objective quality loss for this region, the perceptual quality of the whole video frame is not affected as the objective quality degradation occurs in the non-attention region. The consumed bits for the picture compressed by original HM 11.0 are 499016 bits while the other one are 409640 bits. More details could be found in [8].

5. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we review the progress of perceptual optimization of H.265/HEVC. Specifically, we classify the current perceptual H.265/HEVC video coding technologies into two categories, vision-model based approach and signal-driven approach. Besides the traditional perceptual optimization for the hybrid video coding structure applied on H.265/HEVC, the new feature-oriented perceptual optimization is also designed for the new tools in H.265/HEVC, such as Coding Tree Unit (CTU) structure and Sample Adaptive Offset (SAO). With the better understanding of our HVS, more in-depth investigations for integrating the HVS features and H.265/HEVC are expected. Moreover, although we have wit-

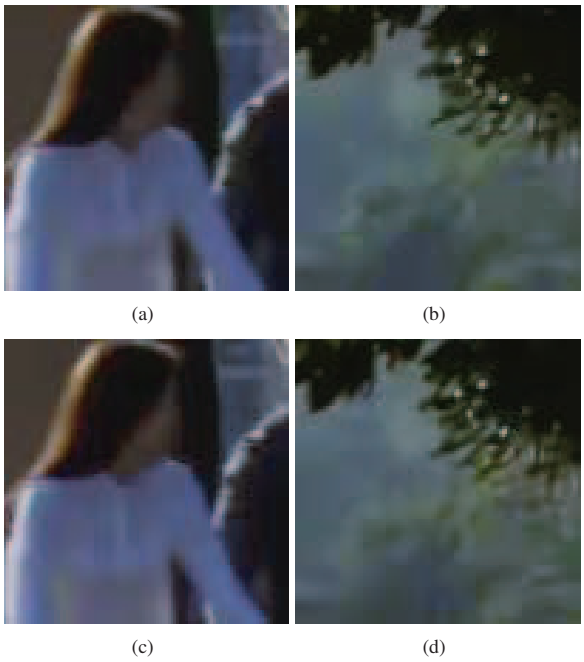

Fig. 3. The coding results of attention region and non-attention region from HM method and attention-based method. Above: HM results; Below: attention-based results.

nessed the attempt to incorporate perceptual quality metric for H.265/HEVC video coding, there are still many aspects in the human perception and their impacts in video representation to be discovered such that advanced video quality metrics could be integrated into the H.265/HEVC system to achieve better visual quality. Further research addressing the multi-disciplinary problems in human vision and signal processing is expected to achieve new breakthrough and make great impact.

6. REFERENCES

[1] G. J. Sullivan, Jens O., W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] S. Li, M. Xu, X. Deng, and Z. Wang, "A novel weight-based URQ scheme for perceptual video coding of conversational video in HEVC," in *2014 IEEE International Conference on Multimedia and Expo*. IEEE, 2014, pp. 1–6.

[3] K. Yang, S. Wan, Y. Gong, H. R. Wu, and Y. Feng, "Perceptual based SAO rate-distortion optimization method with a simplified JND model for H.265/HEVC," *Signal Processing: Image Communication*, vol. 31, pp. 10–24, 2015.

[4] Z. Chen, W. Lin, and K. N. Ngan, "Perceptual video coding: challenges and approaches," in *2010 IEEE International Conference on Multimedia and Expo*. IEEE, 2010, pp. 784–789.

[5] M. Xu, X. Deng, S. Li, and Z. Wang, "Region-of-interest based conversational HEVC coding with hierarchical perception model of face," 2014.

[6] F. Liang, X. Peng, and J. Xu, "Scene-aware perceptual video coding," in *Visual Communications and Image Processing (VCIP), 2013*. IEEE, 2013, pp. 1–6.

[7] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing,*, vol. 13, no. 10, pp. 1304–1318, 2004.

[8] Y. Li, W. Liao, J. Huang, D. He, and Z. Chen, "Saliency based perceptual HEVC," in *2014 IEEE International Conference on Multimedia and Expo Workshops*. IEEE, 2014, pp. 1–5.

[9] . Milani, R. Bernardini, and R. Rinaldo, "A saliency-based rate control for people detection in video," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 2016–2020.

[10] C. Yeo, H. L. Tan, and Y. H. Tan, "SSIM-based adaptive quantization in HEVC," in *2013 IEEE International Conference onAcoustics, Speech and Signal Processing (ICASSP),*. IEEE, 2013, pp. 1690–1694.

[11] A. Rehman and Z. Wang, "Ssim-inspired perceptual video coding for HEVC," in *2012 IEEE International Conference on Multimedia and Expo*. IEEE, 2012, pp. 497–502.

[12] P. Ndjiki-Nya, D. Bull, and T. Wiegand, "Perception-oriented video coding based on texture analysis and synthesis," in *2009 16th IEEE International Conference on Image Processing*. IEEE, 2009, pp. 2273–2276.

[13] J. Kim, S. Bae, and M. Kim, "An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels," 2014.

[14] C. Hoffmann, S. Argyropoulos, A. Raake, and P. Ndjiki-Nya, "Modelling image completion distortions in texture analysis-synthesis coding," in *2013 Picture Coding Symposium (PCS)*, 2013.

[15] V. Adzic, R. A. Cohen, and A. Vetro, "Temporal perceptual coding using a visual acuity model," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014.