HEVC-based Adaptive Quantization for Screen Content by Detecting Low Contrast Edge Regions

Hong Zhang, Oscar C. Au, Yongfang Shi, Xingyu Zhang, Ketan Tang, Yuanfang Guo
Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
Email: {hzhangad, eeau, yshiab, eexyzhang, tkt, eeandyguo}@ust.hk

Abstract— **High-Efficiency Video Coding (HEVC) is the newest video coding standard which can significantly reduce the bit rate by 50% compared with existing standards. The key features and new tools in HEVC are designed for natural video sequences captured by a real camera. Different from natural videos, screen content contain much more edges in text and icon regions. The current video coding standards may blur or even remove low contrast edges, which are very important in screen content for human eyes to recognize the character and the icon. Therefore, this paper proposes an effective modification on HEVC to preserve the low contrast edges in screen content. First, discrete laplacian filter is adopted for edge detection, and then we adaptively adjust QPs for low contrast edge regions, which can be detected based on our designed measurement for edge contrast. Experimental results show that nearly all the regions containing low contrast edges can be detected, and the adjustment of QPs for these regions can greatly protect the edges with no RD performance reduction.**

I. Introduction

Recently, the next generation video coding standard HEVC is being established by ITU-T VCEG and ISO/ICE MPEG organizations. The main goal of HEVC [1] is to significantly improve compression performance, until now it has been designed to achieve equal perceptual video quality using only on average 50% bit rate relative to the performance of previous standards. Based on conventional hybrid coding framework, nearly all the advanced technology in HEVC are developed for natural video content and do not treat artificial screen content specially, which will lead to some undesired performance for screen content.

Screen content refer to images or videos artificially generated by computers, such as word files, PPT slides, and computer games, which are widely used in various applications including remote desktop, game video, etc. Different from natural content, screen content usually contain much more text and icon regions with lots of edges. For human being, these edges are very necessary for recognizing and identifying the text character and icon in screen content. In contrast, natural video content are smooth, and edges in these content will not greatly influence visual perception. Conventional lossy image or video coding standards will reduce high frequencies, and these edges will be blurred. This performance is acceptable for natural content but not for screen content. Therefore, it becomes important to evaluate novel techniques aiming at screen content coding.

Some schemes have been presented for coding compound image which is the combination of text and natural image regions. Considering different texture properties of text and natural regions, researchers propose to do segment on compound image and then treat these two kinds regions separately. In [2] [3], a layer based approach, named mixed raster content (MRC) image model is adopted. They proposed to decompose compound image into different layers. This model requires accurate segmentation for text and natural image parts, and proper filling algorithms for fore- and back- ground layers. People also consider to categorize blocks in compound image into different types depending on their spatial features [4] [5]. In [6] [7], they analyze the properties of different blocks further and develop another two modes RSQ and BICM for H.264/AVC. RSQ mode utilizes the property that signals in spatial domain is sparser for text region than in transform domain, and directly applies quantization to residues without transform. In BCIM mode, blocks are represented by base colors and index map, since colors in text blocks are much fewer than in natural image blocks. The methods mentioned above may be effective for screen content, however, if they are inserted into HEVC much more extra complexity for encoder will be introduced. Besides, the encoded bit stream is incompatible with existing video standards.

In this paper, we take advantage of the state-of-the-art video standard HEVC and propose a non-normative adaptive quantization method for screen content. Since edges are very important in screen content, but treated with less attention in HEVC, we attempt to preserve these edges, especially for low contrast edges as they will be removed by a great chance. Our proposed method is, first, the laplacian filer is applied to screen content frame to detect edges. Second, we propose a simple measurement for edge contrast. Based on this measurement, low contrast edges blocks for certain QP are detected, and then adaptive QPs are computed for these blocks to improve the reconstructed visual performance.

The rest of the paper is organized as follows: In Section II, some key techniques related to our method in HEVC are introduced. Section III presents the proposed adaptive quantization method, including measure of edge contrast, detection for low contrast edge blocks and adaptive QPs selection for low contrast edge blocks. The Experimental results and further discussions are provided in Section IV. At last, a conclusion

978-1-4673-5762-3/13/$31.00 ©2013 IEEE

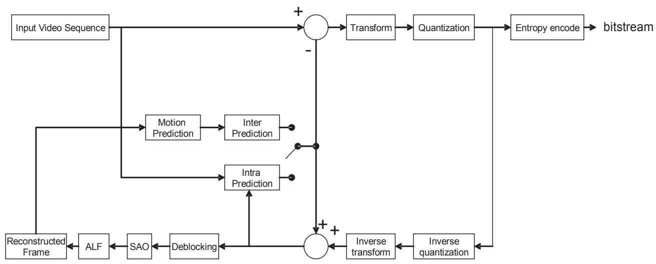


Fig. 1. Hybrid Video Encoder

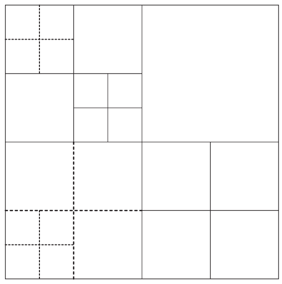


Fig. 2. Partition of CU and TU block. Dash line: CU boundary. Dotted line: TU boundary

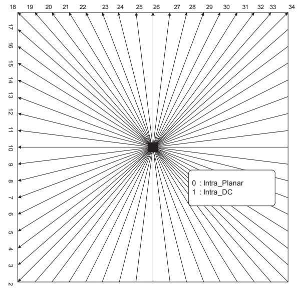


Fig. 3. Modes and directions for intra-prediction

is draw in Section V.

II. HEVC TECHNIQUES

Our proposed method is applied to I slices, so advanced techniques for I slices are introduced. HEVC standard adopts the well-known block based hybrid coding scheme, as depicted in Fig.1. Similar to conventional video coding schemes, the coding performance relies on intra and inter prediction, transform, quantization and effective entropy coder. Besides, block structure in HEVC is more flexible. It defines three new kinds of units, coding unit (CU), prediction unit (PU) and transform unit (TU), and employs quad-tree structures for different unit partitions. CU is the basic intra/inter coding unit, with size from 8×8 to 64×64 for luma component, and can be recursively split into smaller CU size. PU is the basic unit used for prediction mode selection. For intra prediction, PU is the same size as CU except for the smallest CU size where PU can be split further. The root of PU quad-tree is in CU. TU is basic unit for transform and quantization. The size of TU can be from 4×4 to 32×32. Fig. 2 illustrates a possible partition of CU and TU.

In addition, HEVC employs more modes for intra prediction. Compared to 8 directional intra prediction modes of H.264/AVC, there are 35 modes defined for PU, including DC prediction, planar prediction and 33 directional orientations, as illustrated in Fig. 3. For chroma prediction block, mode DC, planar, vertical, horizontal and mode of luma are candidate modes. The new added modes provide lots of directions for prediction which can improve the accuracy of prediction direction and generate rather good prediction result for the current block.

Combined with the use of flexible unit partition structure and various block sizes, HEVC can deal with most unsmooth regions of screen content by trying small block size and many directions to do prediction. However, the edges still can be blurred to some extent for lossy coding. From our experiments, we observe that the reconstructed performance of strong contrast edges in screen content is acceptable, while with the same configuration, the results of the low contrast edges were severely blurred. Therefore, a modification of HEVC becomes necessary to preserve the low contrast edges regions for screen content coding.

III. PROPOSED METHOD

In this paper, we propose to adaptively adjust QP for CU containing low contrast edges. Edge regions should be firstly detected from original picture, and then a measurement of edge contrast is defined. Based on this measurement, regions contain low contrast edges are detected. At last, adaptive quantization is applied to these regions.

A. Edge Detection

There are a lot of ways to perform edge detection, and most of them can be categorized into gradient based method and zero-crossing based method. Gradient methods usually compute the first derivatives G_x of image in x-direction and G_y in y-direction, and search the position of maximum gradient $G = \sqrt{(G_x^2 + G_y^2)}$. The most commonly used kernels are Sobel, Robert and Prewitt operators. Using operator with second derivative of signals for edge detection is called zero-crossing method, and the well-known operator is laplacian filter. The laplacian has the advantage that it is an isotropic measure, as the magnitude of the 2nd derivative is computed in all orientations. In addition, zero-crossing method can detect any positions where intensity gradient tents to increase or decrease, including very low contrast gradient changes. Since there is little noise in computer generated contents, we can take use of laplacian filter to detect all edges for screen contents. We apply the laplacian operator H= $\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ to the original picture, and obtain the edge mask M_H, indicating the positions of edges.

B. Edge Contrast Measure

For conventional hybrid video coding scheme, quantization is applied to the transformed residues and is the main cause of distortion. We hope the contrast measure is directly related to the quantization step, which can help build the relation between the quality of reconstructed block and edge contrast. So, edge contrast measurement C is defined in the frequency domain.

Let F denotes the original picture and $X_{N\times N}$ denotes an $N \times N$ block in the detected edge picture E, where E is equal to $F.{*}M_H$. Based on HEVC, here DCT is adopted to transform $X_{N\times N}$ to frequency domain, and the resulting DCT coefficients are $d_{i,j}, (i, j = 0, 1, ..., N-1)$. Each coefficient $d_{i,j}$ describes the contribution of the corresponding basis function [8], which is arranged in an increasing order of frequency from

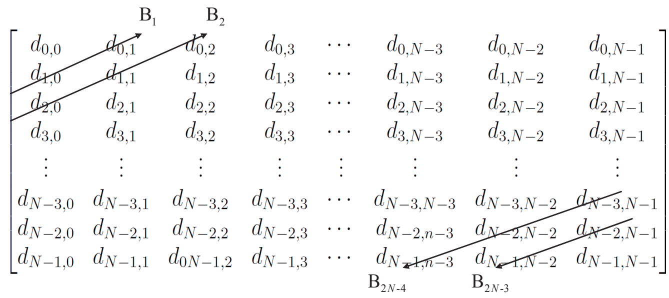


Fig. 4. Classified bands for DCT transform coefficients

left-up to right-bottom. $d_{0,0}$ is DC coefficient and other $d_{i,j}$ are AC coefficients. The spatial intensity changes are properly described by these coefficients in frequency domain, which provides us a natural way to define contrast measurement C in DCT domain.

Transformed coefficients $d_{i,j}, (i, j = 0, 1, ..., N-1)$ are first classified into $2N$-1 frequency bands in a diagonal scan [9]. Fig. 4 illustrates the structure of frequency band, each frequency band $B_t(t = 0, 1, 2, ...2N-2)$ contains coefficients $d_{i,j}$, where $i+j = t$. This is a multiscale structure for different frequency levels, as each band contains the coefficients of basis functions under the same frequency level. Our edge contrast measurement C is designed based on $B_t, (t = 1, 2, ...2N-2)$, except B_0 band.

$$C = \sum_{t=1}^{2N-2} w_t c_t \tag{1}$$

where w_t is the weight for each band B_t and c_t is

$$c_t = \sum_{i+j=t} |d_{i,j}|, t = 1, 2, ...2N-2 \tag{2}$$

which is the sum of coefficients in the t-th band. Edge contrast measurement C is the weighted combination of every AC band contrast measurement c_t, and weight w_t can be used to address interested frequency band. In screen content, we treat all AC bands equally as most text and icon edges are complex, so $w_t = 1/(2N-2)$.

C. Low Contrast Edge Region Detection

Our goal is to improve the performance for low contrast edge regions. It is because that edges are very important for screen content, and HEVC encoder will blur them at the quantization step. Usually, the transform coefficients of picture are smaller for high frequency than for low frequency, after quantization is applied to the transformed coefficients, some AC coefficients will be quantized to zero, which cannot be restored and will cause distortion. Therefore, we expect to search these regions that the edge contrast C drops greatly after quantization, which means most AC coefficients have large distortion and the reconstructed frame is of low quality. Since the quad-tree partitioning structure is complex and time consuming, the detection for low contrast edge regions will not be conducted in every CU partition. Instead, the detection is implemented as a preprocess using a certain block size before current frame is compressed.

First, two-dimensional DCT is applied to block $X_{N\times N}$, and edge contrast measurement C is computed using equation(1), (2). Based on the input $Qstep$, a simple rounding quantization is applied to transformed coefficients

$$q_{i,j} = round(\frac{d_{i,j}}{Qstep}), i, j = 0, 1, 2....N-1 \tag{3}$$

where $q_{i,j}$ is the quantized coefficients. We would like to know the performance of reconstructed result, so inverse quantization is applied,

$$d'_{i,j} = q_{i,j} \cdot Qstep, i, j = 0, 1, 2....N-1 \tag{4}$$

$d'_{i,j}$ is the reconstructed frequency coefficient. Now, we can get edge contrast measurement C' using $d'_{i,j}$ for reconstructed block. If the edge block $X_{N\times N}$ is seriously blurred after quantization, the contrast measurement C' of the reconstructed block will be much smaller than C. It is because most of AC coefficients are small relative to current $Qstep$, and are quantized to zeros which result the reconstructed $d'_{i,j}$ are also zeros. We want to detect this kind block, so a contrast reduction measurement Re is defined

$$Re = 1 - \frac{C'}{C} \tag{5}$$

when Re is larger than a threshold T (T is set $1/3$ in our experiments), we regard this block $L_{N\times N}$ as low contrast region under current QP.

D. Adaptive Quantization

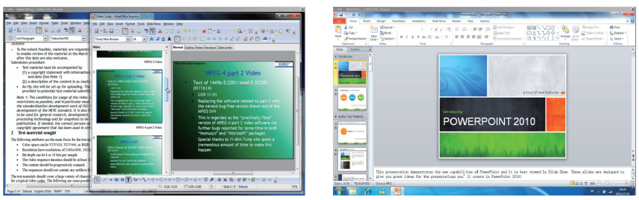
This is a pre-process step, we adjust QP for the detected low contrast edge block $L_{N\times N}$ adaptively. For simplicity, we will iteratively reduce QP by 6, as $Qstep$ will reduce by half when QP decreases by 6. In each iteration, the contrast measure C' and contrast reduction measure Re will be recomputed. The iteration terminates when the measure Re is larger than the threshold T. After all the low contrast blocks are assigned proper QPs, a QP map is build to indicate the QP value for each block in current picture. This map will be used in the encoder for adaptive QP selection and CU partition. It means if one CU contains different QPs, this CU will be split further without compression until the spitted CU covers same QP. With the selected QP for the CU, the following steps are same as HEVC, continue divide current CU if possible, and choose the best partition based on R-D cost.

IV. Experimental Results

We evaluate the performance of proposed method on the reference software HM7.0 [10] of HEVC. Recommended intra-main test conditions by JCT-VC are used, and deblocking filer is disabled as it will blur the decoded screen content. As shown in Fig. 5, two screen content pictures from F class test sequences in HEVC are used as test frames. In our experiments, the block size for detecting low contrast edge regions is set 16 × 16 on Y component, and four base QPs (22, 27, 32, 37) are selected to test our method. We compute adaptive QP for each low contrast edge block according to the base QP.

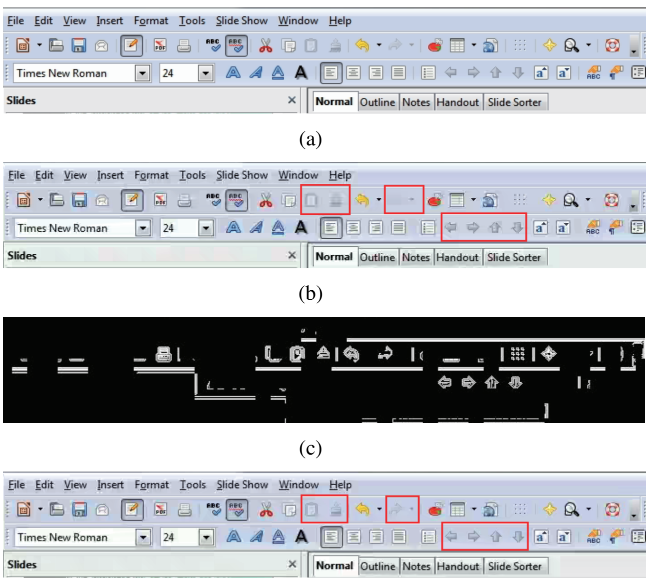
The subjective results are shown in Fig. 6, limited by the length of paper, only part of the result from SlideEditing is presented. Non-black regions in (c) represent detected low contrast edge regions in Y component. Compared with the original picture in (a), we can see that most low contrast edge regions are detected. (b) is the decoded result using HM7.0 when QP is 37. Look at the regions in red rectangles, edges of icons are seriously blurred that makes it difficult to recognize these icons, while other decoded regions are acceptable. Implement the proposed method in HM7.0, the decoded performance of low contrast edge regions are improved greatly due to the adaptive quantization, and result is shown in (d). Edges of icons in red rectangles become more clear, and people can identify these icons easily. The visual quality is improved significantly.

One concern of this method is that R-D performance may decrease. The objective results, bit rate versus PSNR curves of test frames, are depicted in Fig.7. The curves show that the RD performance of our proposed method is almost the same as HEVC (HM7.0), no reduction. Therefore, we can use same bits to encode screen content but obtain better visual performance than HEVC.



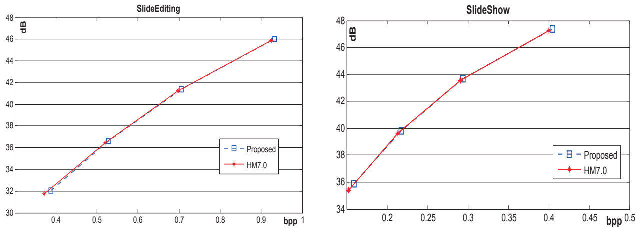
(a) (b)

Fig. 5. Test pictures. (a) SlideEditing_1280x720, (b) SlideShow_1280x720



(a) (b) (c) (d)

Fig. 6. Example Performance of SlideEditing_1280x720, QP is 37 (a) Original, (b) Reconstructed result of HM7.0, (c) Lowcontrast Regions for Y component, (d) Reconstructed result of proposed method



(a) (b)

Fig. 7. Experimental results of rate distortion cost

V. Conclusion

In this paper, we propose an adaptive quantization method to improve the compression performance of low contrast edge regions in screen content. Before compression, a preprocess is conducted to detect low contrast edge regions and build a QP map for adaptive quantization. From the experimental results, we can see that nearly all the low contrast edge regions can be detected under certain QP. After the adaptive quantization is applied to these regions, the decoded subjective performance is good enough for people to recognize these low contrast text and the icons. Compared with the results of HEVC, our proposed method achieves almost same objective RD results as HEVC, however, can greatly improve the subjective compression performance for low contrast regions, which are very important for screen content.

Acknowledgement

This work has been supported in part by the HKUST (HKUST Project no. FSGRF12EG01).

References

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *to apperar in IEEE Trans. on Circuits and Systems for Video Technology*, Dec 2012, pre-publication draft.

[2] L. Bottou, P. Haffner, P. G. Howard, P. Simard, Y. Bengio, and Y. LeCun, "High quality document image compression with "djvu"," *Journal of Electronic Imaging*, vol. 7, no. 3, pp. 410–425, 1998.

[3] R. de Queiroz, R. Buckley, and M. Xu, "Mixed raster content (MRC) model for compound image compression," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3653, no. II, pp. 1106–1117, 1999.

[4] A. Said and A. Drukarev, "Simplified segmentation for compound image compression," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 1, 1999, pp. 229 –233 vol.1.

[5] T. Lin and P. Hao, "Compound image compression for real-time computer screen image transmission," *Image Processing, IEEE Transactions on*, vol. 14, no. 8, pp. 993 –1005, aug. 2005.

[6] C. Lan, F. Wu, and G. Shi, "Compress compound images in H.264/MPEG-4 AVC by fully exploiting spatial correlation," in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, may 2009, pp. 2818 –2821.

[7] C. Lan, G. Shi, and F. Wu, "Compress compound images in H.264/MPGE-4 AVC by exploiting spatial correlation," *Image Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 946 –957, april 2010.

[8] G. Strang, "The discrete cosine transform," *SIAM Review*, vol. 41, no. 1, pp. 135–147, 1999.

[9] J. Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," *Signal Processing Letters, IEEE*, vol. 10, no. 10, pp. 289 –292, oct. 2003.

[10] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-7.0rc2/, Refrence Software HM7.0 of HEVC 2012 [Online].